

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Word Order in Cicero's *Letters to Atticus*
A Multivariate Approach

Kathryn McLachlan

MCLKAT005

A minor dissertation submitted in partial fulfillment of the requirements for the award of the degree of Masters of Arts in Classics.

Faculty of the Humanities

University of Cape Town

2008

COMPULSORY DECLARATION

This work has not been previously submitted in whole, or in part, for the award of any degree. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work, or works, of other people has been attributed, and has been cited and referenced.

Signature: _____

Signed by candidate

Date: 21 May 2008

Signature removed

UNIVERSITY OF CAPE TOWN
GRADUATE SCHOOL IN HUMANITIES

DECLARATION FORM -- MASTERS DEGREE CANDIDATES

Title:	Miss	Student No:	MCLKAT005
Name, Surname:	Kathryn Anne McLachlan		
Postal address: (to which result can be posted)	11 Sierra Court Tiburon CA 94920, USA		
Telephone No's:	1 415 789 5928		
Email address:	kate.mclachlan@gmail.com		
Dissertation Title:	Word Order in Cicero's <i>Letters to Atticus</i> : A Multivariate Approach		
Word count: 19 163 No. of pages 85			

DECLARATIONS:

1. I am presenting this dissertation in **PARTIAL** fulfilment of the requirements for my degree.
2. I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.
3. I hereby grant the University of Cape Town free licence to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever of the above dissertation.

Signature	Signature removed	Date:	21 May 2008
------------------	-------------------	--------------	-------------

IMPORTANT NOTES:

- Candidates for graduation in June and December may expect to receive notification of the outcome of the examination of the dissertation not later than 1st week in June and last week in November, respectively, provided the dissertation was submitted by the due date. Where a dissertation has been submitted well in advance of the due date, earlier notification will be given, if possible. However, the Graduate School in Humanities does not undertake to reach a decision by any specific date.
- Candidates are asked to note that the University will not permit degree/diploma qualifiers to graduate if they have any outstanding fees, fines, interest or dues. **The final date for payment of outstanding amounts is 31 May in the case of qualifiers for June graduation and 31 October in the case of qualifiers for December graduation.**
- Please note that should your examination process run into the following year, you will have to re-register in order to be considered for graduation.

FOR COMPLETION BY FACULTY OFFICE

I acknowledge receipt of three bound and two unbound copies of the Masters dissertation of the above candidate:

Signature		Date:	
Abstract submitted	Yes	No	

University of Cape Town

CONTENTS

INTRODUCTION	1
1. BACKGROUND	3
1.1 LITERATURE REVIEW	3
1.2 WHY CICERO?	17
2. GOLDVARB	27
3. METHODOLOGY	34
3.1 METHODOLOGY	34
3.2 CONSTRUCTIONS	35
3.2.1 ADJECTIVES	35
3.2.2 GENITIVES	36
3.2.3 ADVERBS	37
3.2.4 OBJECTS	37
4. DISCUSSION OF RESULTS	39
4.1 ADJECTIVES	39
4.2 GENITIVES	45
4.3 ADVERBS	49
4.4 OBJECTS	51
CONCLUSION	55
APPENDIX	58
REFERENCES	82
TEXTS	82

Introduction

One of the most striking features of Latin is its flexible word order. Subjects and objects and verbs can be jumbled, seemingly indiscriminately, and there are generally a number of relative or temporal or conditional clauses in the mix as well. Sometimes these sentences can become so long and unwieldy that even their authors have to remind themselves and their audiences what they were saying when they embarked upon them. For example, in *Pro Caelio* 1.1 Cicero elaborates upon the dictates of a law by means of two relative clauses, one embedded within the other, and a tricolon, and then has to start the sentence over, having lost track of where he was grammatically before the subordinate clauses.¹ Small wonder Latin word order has been called a “bugaboo” (Gries, 1951:87) or “unnatural and wholly without plan” (Robbins, 1951:78). However, it is not as random as it appears at first sight. There is a basic order, and the variations upon it are not arbitrary. They are influenced by a combination of factors ranging from syntax to semantics, pragmatics to typology.

In this dissertation, I investigate word order patterns in Cicero’s private letters to his close friend Atticus. My Honours dissertation looked primarily at the arrangement of modifiers and heads within noun phrases in the *Epistulae Ad Atticum* (McLachlan, 2006). This one develops upon my Honours dissertation, and whilst some of the material is perforce the same, such as the literature review, I have increased the number of examples studied for each construction and added to the potential factors influencing word order, as well as examining word order within verb phrases as well. Four constructions are studied in total, two within noun phrases and two within verb phrases. These are (1) adjective and noun order, (2) genitive and noun order, (3) adverb and verb order and (4) object and verb order. Whilst much of the literature that I consulted argues that the main factors influencing word order were communicative or pragmatic (Panhuis, 1982; de Jong in

¹ Idem cum audiat esse *legem* quae de seditiosis scelerisque civibus qui armati senatum obsederint, magistratibus vim attulerint, rem publicam oppugnarint cotidie quaeri iubeat: *legem* non improbet ...

Cicero, *Pro Caelio* 1.1; italics added

“When this man hears that there is a law which orders a daily investigation into those seditious and depraved citizens who, taking up arms, have besieged the Senate, used violence against the magistrates and attacked the Republic: he would find no fault with the law ...”

All translations are my own.

Pinkster, 1983, in Calboli, 1989; Devine and Stephens, 2006), my findings in my Honours dissertation suggested that there were other factors involved, in particular the type or part of speech of the constituent, e.g. whether a genitive was a possessive pronoun or a noun, or an adjective descriptive or demonstrative. In the present dissertation I argue that in general the arrangement of words within a Latin phrase is influenced not just by pragmatics but also by syntactic considerations, although the closer one comes to the sentence level (e.g. the arrangement of objects and verbs relative to each other) the stronger the communicative factors become.

In order to determine the relative strength of the factors, I have used the statistics programme GoldVarb in my study. This has allowed me to produce not only distributional analyses but also multivariate analyses of the data for each of the constructions under study. The latter type of analysis has the advantage of being able to show how the factors combine to influence word order – which of them are truly independent of each other, which ones interact, which of them are important and which of them have little or no effect at all. Further, when selecting the significant factor groups it ranks them in order of strength or importance relative to each other, all of which leads to a better understanding of the data and the contextual factors, be they syntactic or pragmatic or even prosodic, that result in variant word orders.

CHAPTER 1

Background

1.1 Literature Review

Like other ancient Indo-European languages, such as Sanskrit, Hittite and Classical Greek, Latin has a system of inflections to mark the grammatical rôle of constituents within the sentence, which enables it to have flexible word order. So the sentence *Incitatus Caligulae equus vinum bibit* “Caligula’s horse Incitatus drinks wine” means the same as *vinum bibit Incitatus equus Caligulae*, *bibit Incitatus vinum equus Caligulae*, *equus Caligulae bibit vinum Incitatus*, and any other possible combination of these words.¹ The relationships between the constituents of the sentence are marked by their endings, rather than, as in English, by their positions relative to each other:

Incitat.us Caligul.ae equ.us vin.um bib.it
Incitatus.nom Caligula.gen horse.nom wine.acc drink.3sg present active

It is obvious, then, that word order has the potential to be jumbled without loss of sense, since all the important syntactic rôles are encoded within the inflections.

That is not to say, as Thomas Arnold, the nineteenth century classicist and grammarian, observes, that there are no “general principles” of Latin word order (1938:17). In his *Latin Prose Composition*, Arnold lays out basic orders for the writing of Latin prose, instructing pupils to note departures from these norms when reading Latin texts. So, in unmarked sentences, word order is subject-initial, verb-final, although if the verb is emphatic it shifts to the beginning of the sentence and the subject is correspondingly moved to the end, these being the two most prominent positions.² Direct and indirect objects precede the verb, genitives follow their governing noun, unless the noun is

¹ Although some combinations will give an ambiguous reading e.g. *vinum Caligulae equus Incitatus bibit* could also mean “The horse Incitatus drank Caligula’s wine”.

² The two emphatic positions in a Latin sentence are the *beginning* and the *end*. By the former our attention is raised and suspended, while the full meaning of the sentence is rarely completed till that last word is reached.

Hence, from the habit of placing the most important part of the predicate, which is generally the verb, last of all, we rarely see a Latin sentence from which the last word or words can be removed *without destroying the life*, so to speak, of the whole sentence.

Arnold, *op. cit.*, p. 18, italics original

preceded by an attributive adjective, in which case genitives are interposed between adjective and noun. Adjectives tend to follow the nouns they modify, unless they denote size or quantity, or the noun they agree with is the head noun of a genitive expression. Numerals and demonstrative and interrogative pronouns precede the nouns they qualify. Phrases in apposition come directly after their head nouns, whilst adverbs and adverbial phrases precede the verbs they modify. Adverbs of time tend to supplant the subject in sentence-initial positions, whilst negative adverbs are always placed directly before the word they qualify (*ibid.*, pp. 17ff.). However, as Arnold points out, there are always exceptions to these rules, “[f]or the order in Latin is determined not only by general principles, but also by considerations of emphasis, clearness, sound, rhythm, and variety” (*ibid.*, p. 22).

Because it is so elastic, Latin word order has often been felt to be unrestricted by any grammatical rules – e.g. Panhuis (1984, quoted by Elerick, in Calboli, 1989:570) states that “from a syntactic point of view, word order in Latin is indeed almost free.” However, that is not to say that there is a complete absence of coherence in the arrangement of the words: “pragmatically it is not free, but organized in a communicative perspective” (*ibid.*). Panhuis’s approach to Latin word order is accordingly focused not on rhetorical patterns, nor on syntax, but on the rôle played by communication (1982:1).

Panhuis’s (1982) study of Latin word order within the sentence is based on the works of Caesar and Plautus, as both authors largely shun literary ostentation (*ibid.*, p. 5), the former copying the “sober style of the Roman annalists” (*ibid.*) described by Cicero (*De Or.* 2.54) as being “non exornatores rerum”³, the latter writing in the “sermo cotidianus”⁴ (*Cic. Or.* 20.67). They are also, respectively, exemplars of Classical Latin – as Posner (1966:30) calls it, the “arty”, literary language – and colloquial Latin. The basis for Panhuis’s investigation into the communicational relevance of sentence constituents was the Functional Sentence Perspective (FSP) of Firbas.

The Functional Sentence Perspective examines the ways in which sentences are structured according to the information they impart. Elements of the sentence which carry information “already known ... or taken for granted” (Firbas, 1957, quoted by Panhuis,

³ “Not given to embellishing their language.”

⁴ “Everyday speech.”

1982:10) have less communicative dynamism (CD), i.e. they do not convey as much new information as other elements may do. An element with low CD is called the *theme*, whilst one with high CD is known as the *rheme*. Matthews (1997:377) defines the theme as that part of the sentence which coincides with what the sentence in its entirety is about, whilst the rheme elaborates upon the theme by providing new information (*ibid.*, p. 321). In general, if all else is equal and there are no other overriding grammatical rules (as in English, where word order is required to make syntactic sense), the theme, with the lowest degree of CD, will come earlier in the sentence than the rheme, with the highest degree of CD (*ibid.*, p.139).

Panhuis found that the standard word order in Plautus's colloquial Latin is theme-rheme, whilst the emotive, or marked, order is rheme-theme (1982, p. 54). He goes on to call "the FSP ... the determining principle of word order" (*ibid.*), with syntactic patterns explicable as "fossilized instances of the communicative word order principle" (*ibid.*, p. 59). For example, the relative pronoun is invariably positioned at the head of the relative clause because as an anaphoric constituent (or in some cases cataphoric), referring to an antecedent in a separate clause, it is dependent on its context and so is thematic (*ibid.*, p. 60).⁵

De Jong (in Calboli, *op. cit.*, pp. 521-40) approaches Latin word order from a similar angle, explaining the position of the Latin subject – unmarked initial or marked non-initial – in terms of whether its function carries priority or posteriority.⁶ So, for example, if the subject has the function of focus, it contains new information, i.e. it forms the rheme of the sentence, and is shifted towards the end, into a non-initial position (*ibid.*, p. 531). The placement of the subject is not just dependent upon the rôle of the subject itself, though. Other constituents can move to the start of the sentence, shunting the subject from its initial position. Anaphora (*ibid.*, p. 524) is the most common of these priority factors, because, as noted above, an anaphor, depending upon given information provided by the sentence, is thematic. Contrastive or emphatic constituents also tend to

⁵ Panhuis does come in for some criticism from Del Vecchio, who suggests that "in some cases, the attribution of a thematic or rhematic character to a constituent is not wholly convincing and instead seems motivated by the necessity of making all the examples fall under the theory assumed" (in Calboli, *op. cit.*, p. 541).

⁶ Priority corresponds roughly to low communicative dynamism in the FSP, posteriority to high.

shift forward within their clauses, presumably because they too are related to context.⁷ Contrast relies upon the presence of some other comparable but different element, otherwise it cannot exist: it presupposes information already given. Emphasis shows that something is more important or prominent than something else, which may not be explicitly stated, merely understood. It is thus a form of contrast.

Elements which act as the setting for the new information contained in the sentence, such as adverbs and adverb clauses, may move forward as well, again in accordance with the principles of FSP. The last priority factor is topic, namely that part of the sentence which identifies the elements about which new information is presented (*ibid.*, p. 523) – topic bears obvious parallels to Matthews's definition of theme (see above). Conversely, de Jong observes that topic can also carry posteriority and shift a subject backwards from the initial unmarked position (*ibid.*, p. 535). He states that this anomaly occurs only within a specific context, where subjects are “more or less predictable”, “added merely for the sake of clarity and ... only one step away from zero-pronominalisation” (*ibid.*).

Finally, de Jong defends his original assumption that the subject is the unmarked position, explaining that while all the non-initial subjects can be explained in light of priority or posteriority factors, there are some initial subjects that have no motivation (*op. cit.*, p. 536). Because they cannot be explained, he assumes this to be their unmarked position. He also cites the example of the sentence “pons lapideus flumini impositus iungit urbem” (Curt. 5.1.29),⁸ which has a focus constituent for a subject, carrying new information and posteriority, as well as a topic element which has a priority. However, the subject is still retained in sentence-initial position: this must therefore be the unmarked position for the subject, since it overrides both priority and posteriority (*ibid.*, p. 537). Subject-initial word order is also one of Arnold's general principles (see above).

Pinkster (1990:171f.), though, observes that subjects are often not explicitly realised as separate elements in Latin, since they are already indicated by the inflection of the verb. The first and second person are instantly identifiable for speaker and addressee, for obvious reasons, and do not need to be expressed unless for emphatic purposes. A third person subject, who is not a participant in the conversation, will again be reflected by the

⁷ The “principle of context” (Panhuis, 1982:55) is one of the factors making up the FSP, along with the natural distribution of CD and the semantic domain (*ibid.*, pp. 10-11).

⁸ “A stone bridge built over the river joins the two halves of the city.”

verb's ending, but the specific reference will not be. However, Pinkster presumes that the identity of the third person can be deduced from the context – prior mention as some other constituent a little earlier, for example – and so does not necessarily need to be explicitly expressed. Further, in an analysis of Cicero *Att.* 1.5, he demonstrates that out of five subjects, only three are placed at the very beginning of their sentences. He then proceeds to analyse the constituents that are found in this position: part of an embedded clause, relative and interrogative pronouns (which are Topic and Focus respectively), subordinate clauses and preposition phrases that present the new subject and so state what the sentence will be about (Theme). Naturally, connecting particles such as *itaque*, *nam*, *et* and *sed* will also take first place, joining one sentence with what has gone before. So, instead of the initial position being reserved for the subject, as de Jong argues above, Pinkster reserves it for Topic and Focus, which may be the subject, but equally may not be.⁹ Similarly, Devine and Stephens (2006:87; see below) suggest that the subject originates within the verb phrase, but is then raised to sentence-initial position. This may also account for sentence-final subjects, or tail subjects, which do not have focus, but rather contain old information.

There are, of course, other methods of studying Latin word order, such as the syntax-based approach of Elerick (in Calboli, *op. cit.*, pp. 559-91). In this paper, Elerick demonstrates that Latin has a dominant verb-final order,¹⁰ not by comparing the frequency of occurrence of verb-final and other orders, as has frequently been the line of approach, but through an examination of Latin gapping in Cicero's orations, i.e. the

⁹ Cf. also Walker (1918:2f.):

I do not believe that Caesar and Cicero knew that the subject should stand first. The rule may be the best that we can do for pupils who are asked to translate into Latin wholly detached sentences from a composition book; but I believe that Caesar and Cicero, if requested to translate the same sentences, would ask what came before them. Neither Caesar nor Cicero has left us a single completely detached sentence; that is, a sentence not connected with some other thought already expressed or in his mind ... In fact, if one will check off with pencil and paper all the sentences in a few pages of Caesar ... he will find that the great majority begin with whatever word links the thought of the sentence with that of the preceding sentence, whether the linking word is subject or ablative or verb, whether it is emphatic or not.

¹⁰ This is one of Arnold's general principles. See above. Pinkster (1990:88) points out that although SOV predominates in both Caesar's commentaries and Cicero's *Epistulae Ad Atticum*, it is by no means universal. At the same time, Plautus's Latin, dating from somewhat more than a century earlier, tends towards SVO, but again it is not fixed. As a result, he concludes, there are no grounds for stating that Latin favours one word order over another until further research has been done.

omission of the verb in all but one of two or more co-ordinated clauses, where the deleted verb can be recovered from the “manifested instance” (*ibid.*, p. 562). Elerick also discusses Noun Phrase (NP) gapping, which occurs when a deleted NP is part of a sequence dominated by a different subject from the subject dominating the sequence in which the NP appears (*ibid.*, p. 564). If the deleted constituent forms part of the same sequence in which the manifested constituent appears, however, this is not gapping, but together with gapping falls under the heading of “coordinate deletion”, to use Seligson and Knudsvig’s term (1974, quoted by Elerick, *op. cit.*, p. 566).

To return to the point in hand, gapping, Elerick’s argument runs, is marked, and therefore more difficult to parse, and, as a result, it discourages the presence of other marked constructions, such as a deviation from the standard word order. This he calls Preemptive Markedness (PM), which can be used to test the dominant word order of Latin by establishing a basic order used in those sentences which contain elements structured in such a way to be construed as marked (*ibid.*, pp. 561f.).

If Latin is an SOV language (cf. note 10), when a verb is gapped, it will be gapped leftwards, i.e. SOV + SOV > SO + SOV. On the other hand, if it has SVO as its dominant word order, it will gap rightwards, i.e. SVO + SVO > SVO + SO¹¹ (Ross, 1970, cited by Elerick, *op. cit.*, p. 566). If Latin has a free word order, we can expect it to behave like Quechua and Cherokee, which are “extremely free”, and gap rightwards with variant orders SOV + SO and SVO + SO (Pulte, 1973, cited by Elerick, *op. cit.*, p. 570). The examples provided by Elerick in the article all gap leftwards, e.g.

Sullam res publica (revocavit)¹², Murenam Sulla revocavit¹³
 O S V O S V

insignia (reportarent), victoriam reportarent¹⁴
 O V O V

Elerick, *op. cit.*, p. 569

In other words, it would appear that Latin’s basic word order is OV, as this is the default arrangement of constituents when gapping – an example of Preemptive Markedness –

¹¹ This is the typical English gapping word order, e.g. “I ate an apple and Jennifer ate sushi” becomes “I ate an apple and Jennifer sushi”.

¹² Brackets indicate a constituent not present in the actual Latin quote, but supplied by Elerick to show which constituent has been gapped.

¹³ The republic recalled Sulla, Sulla recalled Murena.

¹⁴ They brought back decorations, they brought back victory.

occurs. Elerick also notes a chiasmic pattern, i.e. X (Y) (X) Y, to which coordinate deletion is “apparently constrained” in cases where two different constituents are deleted (*ibid.*, p. 565), and furnishes further examples:

maiores consuetudine (paruisse) (maiores) utilitate paruisse¹⁵
 S O V S O V

ibid., p. 568

Thus Elerick’s findings are in contention with Panhuis’s claim that syntactically, Latin word order is not governed by any rules (1984, quoted by Elerick, *op. cit.*, p. 570).

Having covered the general question of Latin word order at sentence level, we now turn to the specific question of the ordering of the constituents within noun phrases. De Jong (in Pinkster, 1983, pp. 131-41) discusses several factors that determine the position relative to their head nouns of modifiers referring to persons, i.e. “possessive pronouns, and genitives of nouns, proper nouns, and demonstrative and anaphorical pronouns” (*ibid.*, p. 131). Although the postposed modifier is recognised as representing the unmarked order, preposed modifiers occur roughly as frequently, being found when they have the function of contrast, topic,¹⁶ referential unity or context reference. Contrast is of particular importance (*ibid.*, p. 132), since a contrastive modifier specifies the difference between its head noun and whatever it is compared with. As a result it is only natural that it should be brought into greater prominence. For example, in “*e suis finibus in Helvetiorum fines*” (Caes. *Gall.* 5.29.3, italics de Jong’s),¹⁷ the NPs share the same head nouns but differ in their reference as specified by the modifier (*ibid.*, pp. 132 f.). Intrinsic contrast is the reason given for the positioning of the normally postposed possessive adjective before its head noun, as in “*quot ego tuas petitiones ... effugi?*” (Cic. *Cat.* 1.15, italics de Jong’s),¹⁸ where the adjective highlights the relationship between Cicero (“ego”) and Catilina (“tuas”) and contrasts their actions (*ibid.*, pp. 134 f.).

Topic also causes the modifier to shift forward, often leading to hyperbaton, and may be combined with contrast. It provides the setting for and presents the head noun, which will be elaborated upon in the rest of the sentence (*ibid.*, pp. 135 f.). When the personal modifier and head noun refer to the same class of people, or when the modifier refers to

¹⁵ “The elders complied with custom and with expediency.”

¹⁶ Both of which functions he describes as carrying priority in De Jong (1989).

¹⁷ “From their own lands into the lands of the Helvetii.”

¹⁸ “How many of your attacks ... did I escape?”

the parts of which a collective or plural head noun is made up,¹⁹ the modifier is again preposed (*ibid.*, p. 137). Finally, an anaphoric modifier – one with “context reference” – will be preposed as well, presumably because it helps to specify the head noun in terms of some fact already known (*ibid.*, p. 138). These four functions of the personal modifier, of course, can be compared to the functions discussed above that influenced subject position in the sentence (de Jong, in Calboli, *op. cit.*).

Elerick (in Herman, 1991, pp. 67-74) discusses the phenomenon of word order doubling and its implications for Latin, in particular in NPs. Doubling refers to the “co-existence of alternate orders for the constituents of a syntactic string” (*ibid.*, p. 67), and it is thanks to its prevalence in Latin, Elerick argues, that its word order is considered to be free. Here he makes use of Vennemann's Natural Serialisation Principle (NSP), which states that the arrangement of verbs and their complements tends to be paralleled in other constructions where elements are in the same operator-operand relationship (Matthews, *op. cit.*, p. 240). The chief use of the NSP, for Elerick, is its ability to define harmonic and non-harmonic word orders – one of the co-existing doubled orders will thus be harmonic, falling in with the verb-complement order in accordance with the NSP, and the other non-harmonic (1991:67). He also introduces the principle of Harmonic Phenotypic Linearization (HPL), which examines whether the ordering of a group of constituents is harmonic or not (*ibid.*, p. 68).

Elerick, assuming from the basis of his earlier work (in Calboli, *op. cit.*, above), that Latin has OV order, gives four different examples of the HPL in action. The first one is the distribution of OV and VO gerundive orders, first with a nominal head, and second with adpositions. For the nominal plus gerundive construction, there are four orders, given below with the number of instances of each in Elerick's corpus:

- | | |
|----------|----|
| 1. [OV]N | 19 |
| 2. N[OV] | 19 |
| 3. N[VO] | 14 |
| 4. [VO]N | 1 |

1 and 2 both have the dominant OV word order, which would appear to override the non-harmonic order of the head noun in 2, whilst 3, although it has the non-dominant VO

¹⁹ Partitive genitives fall into this category – provided that they are co-referential with their head noun. Hence the postposed partitive genitive with *pars*, *quisque* and other such expressions (de Jong, 1983:137).

word order, has a harmonic head noun and occurs more frequently than 4, which has non-dominant word order and a non-harmonic head noun. Similar results are produced for the adpositions, with dominant word order and a harmonic adposition increasing the number of times a constituent occurs (*ibid.*, pp. 68f.).

A second example of the HPL in action is that of the distributions of the order of the possessive adjectives with and without prepositions. Again, harmonic orders prevail over non-harmonic orders, and Elerick notes that the difference between the harmonic construction $p + NAdj$ and the non-harmonic $p + AdjN$ is proportionally greater than the difference between the two word orders “when there is no preposition to trigger HPL” (*ibid.*, p. 69).

Similar results are obtained in the case of genitive-head noun (GN) orders with and without prepositions. However, Elerick takes his material from two different authors, namely Caesar from the first century B.C. and Livy, who wrote from the late first century B.C. to the early years of the first century A.D., and compares the distributions for both. Here, the point of interest is that although both results show the bias towards harmonic linearization, in Livy that bias is less obvious than it is in Caesar. Elerick argues that this is the result of language change, with head noun-genitive (NG) becoming the dominant word order and GN, originally dominant, although co-dominant with NG in Classical Latin, dropping out in early Romance, and thus that the HPL can be used to track language change as the bias begins to level out and then tilts in the opposite direction (*ibid.*, pp. 69ff.).

The final doubling that Elerick examines is that of adjective-noun (AdjN) order, when the noun is a genitive and part of a genitive-noun construction. His examples are all taken from Cicero, a Classical author, so it is not too surprising that NG and GN are pretty much on a par, whilst AdjN is clearly preferred to NAdj. As with the other examples cited above, the non-dominant, non-harmonic order $[GAdj]N$ occurs least often – it appears less than half as often as the harmonic $N[AdjG]$, which has dominant AdjN word order, for example (*ibid.*, p. 70). So the HPL provides a means of determining which word doubling patterns are more prevalent, as well as explaining and predicting them in terms of a combination of harmony and dominance. It can also be used to track and explain language change.

Devine and Stephens (2006) provide a comprehensive discussion of Latin word order with their “multidimensional analysis” (*ibid.*, p. 6) based on the syntactic, semantic and pragmatic interpretation of evidence amassed from a variety of authors and texts. However, their aim is not “broad coverage of the subject matter”, but instead the clarification of “theoretically significant issues” (*ibid.*, p. 8), and so their approach generally involves close analysis of the patterns associated with “just a few representative words” (*ibid.*), followed by an integration of these patterns into a theoretical framework.

From a “dense philological analysis” of the arguments of verbs (*ibid.*, p. 79), including subjects, objects, locatives, instruments, adverbial adjuncts and ablative absolutes, they propose the following neutral word order for simple sentences in Latin as follows:

Subj DO IO/Obl Adj Goal/Source Nonref-DO V
or in English: Subject – Direct Object – Indirect Object or Oblique argument –
Adjunct – Goal or Source argument – Nonreferential Direct Object – Verb

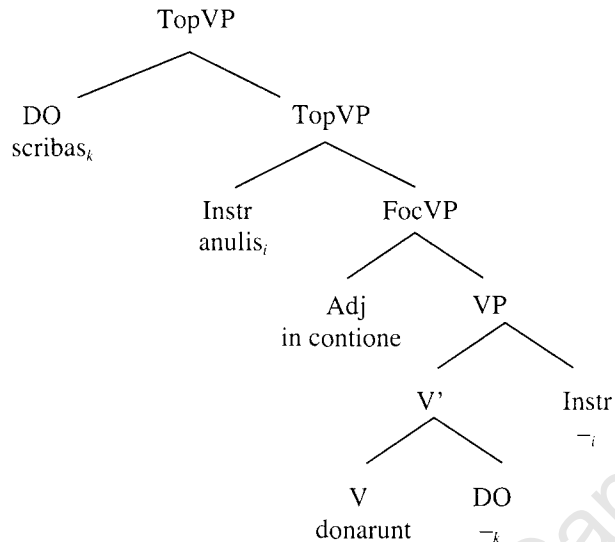
ibid.

Precisely because Latin has a neutral word order, and because this order is stable and consequently fixed to a certain degree, Devine and Stephens argue that there must be an underlying hierarchical structure for the Latin sentence (*ibid.*, p. 83), and suggest that specifier syntax (as in X-bar theory), with attendant argument raising, is the most efficient way of diagramming Classical Latin sentences (*ibid.*, pp. 87ff.). So the arguments of the verb – the direct and indirect objects as well as the oblique arguments and instrumental adjuncts – originate as complements following the verb, which are then moved to higher specifier projections from which they c-command²⁰ the verb phrase (VP). Time and place adjuncts, however, may be generated in a VP-external position, depending on their function (*ibid.*, p. 92). The projections into which the arguments raise are named for their pragmatic function, the leftward projections being called TopVP, whilst the rightmost projection is FocVP, where Top and Foc stand for “topic” and “focus” respectively.²¹ On the other hand, the complement positions from which the arguments raise are named for their grammatical functions of direct and indirect object

²⁰ One constituent in a phrase may c-command, or constituent command another, when the first branching node that dominates the first constituent also dominates the second constituent, cf. Radford (1997:255).

²¹ Cf. de Jong (in Calboli, *op. cit.*) above for the topic position being sentence-initial and the focus sentence-final.

(*ibid.*, p. 87). So they diagram “scribas anulīs in contione donarunt” (Cic. *Verr.* 2.3.185)²² as follows (p. 88):



In general, according to Devine and Stephens, the complements raise in antimirror order, so that if they originate after the verb in the sequence DO, IO, Instr, they will be found in the same order in the surface structure (*ibid.*, p. 92). The only exceptions to this rule are directional phrases, which raise to the FocVP, and nonreferential objects, especially those that are part of set phrases such as *gratias agere* “to give thanks” or *iter facere* “to make a journey”, which raise to a specifier position between directional phrases and the verb (*ibid.*, pp. 93f.).

Irregular orders which violate the neutral organisation of verb arguments can be attributed to scrambling, which in Latin is “typically not a permutation of arguments inside the verb phrase [but] the movement of an argument from inside to outside the verb phrase” (*ibid.*, p. 108). From the tree diagram above, it can be seen that in unmarked word order the arguments do not raise outside the verb phrase, since the highest projection remains a VP; when scrambled, they move even further leftwards and out of the VP. This can be proved by means of the adverb test, using adverbs of manner (*ibid.*, p. 99). Essentially, arguments which appear to the left of an adverb of manner are scrambled, since they have to move over a constituent which c-commands the verb, or at

²² “They gave their scribes rings in the assembly.”

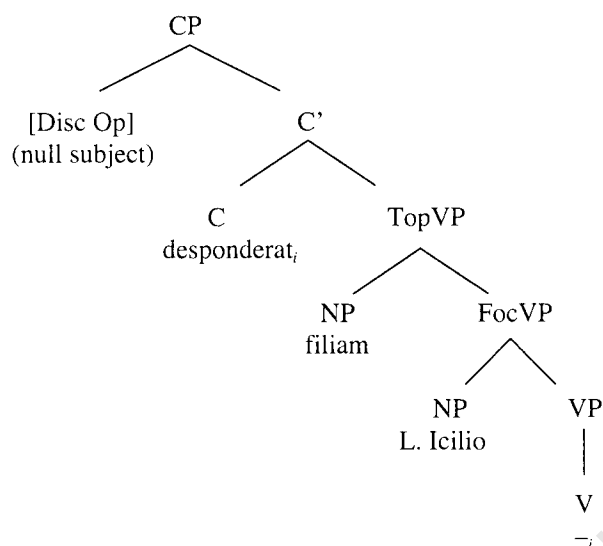
times even the verb phrase as a whole, and thus out of the VP. Devine and Stephens also argue that the adverb test proves the existence of string vacuous scrambling, i.e. the neutral word order remains unchanged, even though certain elements have been scrambled, or moved out of the VP (*ibid.*, p. 108). The arguments are still arranged in their antimirror order, but without the adverb to mark the VP boundary we cannot tell if one or more of them is outside the verb phrase.

VO order, though, poses a potential problem for specifier syntax, which requires that all post-verbal complements raise to specifier positions. To get around this, Devine and Stephens suggest that Latin speakers had the option of switching to V-bar syntax for these “VO leakages” (*ibid.*, p. 127), which are comprised of tail objects with weakened prosodic stress and the pragmatic property of topic²³ as well as “abstract nouns in ... fixed phrases” (*ibid.*, p. 135). For both of these types “the event is presented without individuation of the object” (*ibid.*), which is to say that the object is either not terribly important to the action or event expressed by the sentence (the tail objects) or, in the case of the abstract nouns, they do not really refer to anything, being part of a rote expression.

Sentences with verb-initial orders, by contrast, are not the result of mass argument stranding, but rather of the movement of the verb to the higher position of the complementiser phrase (CP), as can be seen from scrambling amongst the arguments that now follow the verb. This is the result of the adoption of athetic perspective in the sentence. In athetic statement, the subject's reference is not fully complete until the sentence itself is complete, e.g. “This is my horse Amber”. In categorical statements, on the other hand, the subject already has a known reference and the event described in the sentence is a property or state of affairs predicated of it, e.g. “Amber jumped the fence” (Kurzová, 1993:84). Devine and Stephens suggest thatthetic sentences generate a higher projection, generally a CP, whose specifier position is filled by a null subject, which is either a “discourse or context bound subject”. The null subject then compels the verb to move into the head position of the CP as in “desponderat filiam L. Icilio” (Livy 3.44.3)²⁴ below (p. 168):

²³ Cf. de Jong (in Calboli, *op. cit.*, p. 535) above on sentence-final subjects.

²⁴ He betrothed his daughter to L. Icilius.



Verb-second order – that is, when the verb follows one of its arguments but precedes the rest – also involves verb raising. This time, the preceding argument carries the function of topic, whilst the following argument has the function of focus. In the sentence tree showing the structure of the neutral word order, the focus position is the projection immediately above the VP. Thus the focused argument raises into the specifier position of FocVP, which leads to the topical argument and the verb moving to the specifier and head position of TopVP (*ibid.*, p. 179).

One of the constructions analysed by Devine and Stephens, and which is also investigated in the present study, concerns nominal arguments, in particular genitive constructions and adjectives. Regarding the genitives, Devine and Stephens claim that these originate in a complement position, following the head noun that they modify (*ibid.*, p. 378), with prenominal genitives undergoing raising from their default position when carrying focus. They propose three different syntactic structures for the genitive construction – the minimal structure theory, the maximal structure theory and the functional theory – as well as providing pragmatic and prosodic motivations for the pre- and posthead orders of the genitive.

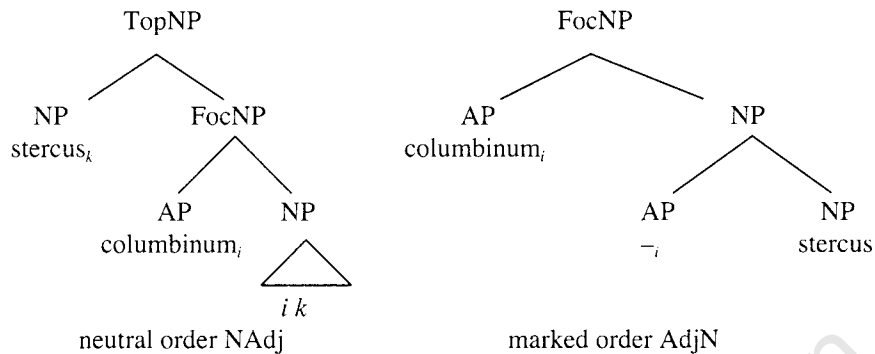
The minimal theory relies on X-bar syntax, with the NP providing just the three positions of specifier, head N and complement, whereas the maximal theory creates a multiplicity

of focus and topic positions (*ibid.*, pp. 380ff.). However, as Devine and Stephens observe, the minimal theory, with only three positions available, cannot account for pragmatic considerations involved in genitive raising, whilst the maximal theory faces the problem of projection overload. Moreover it is also simpler to generate the marked order GenN than it is to generate the neutral NGen (*ibid.*, p. 384). The functional theory attempts to solve these problems, retaining the pragmatic articulation of the maximal theory whilst cutting down on the number of potential structures available by ensuring that the base NP is at the very least partly evacuated.

Aside from the considerations of topic and focus, another possible motivation for genitive raising has to do with “conceptual individuation” (*ibid.*, p. 388) – does the genitive construction express one single concept or are the genitive and noun two separate concepts? Premodifier genitives would seem to occur when the construction is felt to express one single concept – the example cited is *mortis metus* “fear of death” – while postmodifier genitives are individual, independent items of information. Further, if the relationship between possessor and possessum is predictable, the genitive case possessor is more likely to be a premodifier (*ibid.*, p. 389). For example, the relationship “son/daughter of” is more predictable than “uncle of”, and so we find *Quinti filiam* “daughter of Quintus” but *avunculus adolescentis* “uncle of the young man” (*ibid.*, pp. 352ff.)

The prosodic motivation of genitive raising is based upon pragmatic factors. When strong focus falls upon the genitive in its original complement position, this leads to an iambic rhythm which conflicts with the natural trochaic cadences of Latin. As a result, the genitive moves across the noun and so the stress pattern reverts to a trochee (*ibid.*, pp. 391ff.).

Devine and Stephens also analyse the arrangement of adjectives and nouns within Latin NPs in terms of the minimal and functional structures. For the minimal theory, the underlying order is NAdj, as opposed to AdjN for the functional theory, which Devine and Stephens argue is “superior” to both the minimal and maximal structure theories for representations of modifiers (*ibid.*, p. 484). As with the genitives, in the functional analysis the neutral NAdj order is once again more complex than the marked order (*ibid.*, p. 483f.):



Adjective raising, according to Devine and Stephens, is triggered by the type of adjective – to move, it must be intensional (e.g. *vetus* in the sense of “previous”), scalar (e.g. *magnus*, *parvus*), or nonliteral (e.g. *superior* in the phrase *superior dies*, versus *superior* in *locus superior*). These adjectives must be closely linked with their nouns to be interpreted correctly; however, in the neutral order, the noun is virtually scrambled out of the NP into the topic position, becoming “semantically independent” of the rest of the phrase, and so it is harder for adjectives to “access intensional components in the meaning of the noun” (*ibid.*, p. 485). If the noun does not raise, however, but the adjective still does, the adjective can c-command the NP from the FocNP.

So the literature on Latin word order does not point to an arbitrary collocation of constituents, but instead suggests that there are underlying patterns, whether they are based on communicative practice or on syntactic rules or even – as seems most probable – both. After all, even if Latin has a basic word order, its substantial inflectional system allows it to transpose constituents for emphasis or aesthetics²⁵ without violating the sense or the syntactic coherence of a sentence.

1.2 Why Cicero?

In his *Aspects of the Theory of Syntax*, Chomsky defines the purpose of linguistics:

Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory

²⁵ See below.

limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance ... Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behaviour. Observed use of language or hypothesized dispositions to respond, habits, and so on, may provide evidence as to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics ...

1965:3f.

In other words, the study of language is not concerned with language as it is used in everyday life by real people, a view which has understandably resulted in a great deal of controversy. In order to establish the underlying syntactic rules of a language, it is necessary to rely on the “intuitive knowledge” of its speakers (*ibid.*, p. 8), or their linguistic competence, rather than on their linguistic performance, error-riddled as it is. When it comes to Classical Latin, however, it is simply not possible to conduct linguistic study from this viewpoint, because there are no speakers of Classical Latin whose grammatical intuitions we may rely upon to formulate theories of Latin grammar. Instead, all our evidence for Latin comes from texts – the written product of real speakers. In order to study Latin grammar, we have to study the words of real people, with all their flaws and inconsistencies.

So it may seem that the corpus-based study of Latin perhaps has more in common with the sociolinguistic approach to language, which relies upon “[e]mpirical [f]oundations” (Labov, 1972:xvi). However, sociolinguists tend to avoid texts as sources of information about language, since for them the project of linguistics is precisely the actual use of language by real speakers that Chomsky rejects. This cannot be gleaned from “the inevitable obscurity of texts, the self-consciousness of formal elicitation, and the self-deception of introspection” (*ibid.*, p. xiii), but rather through recording normal interactions in ordinary social conditions. The target of this sort of investigation is the vernacular, the style of speech used by people when they are most relaxed and least self-conscious, and so the most natural form of speech. Written texts, on the other hand, are anything but natural. They allow their author time for thought and the freedom to rework them for maximum effect – the written word can hardly be described as anything but self-conscious. Moreover, the surviving Greek and Latin texts that we have are products of a tradition steeped in rhetoric, the art of manipulating language.

Rhetoric first found a place in ancient curricula in the fifth century B.C., with the rise of the democracies in Greece. In Athens, for instance, success in politics was increasingly dependent upon the ability to sway the Assembly through the power of rhetoric – Guthrie (1971:19) cites the Athenian reaction to the revolt of Mitylene as an example of the susceptibility of the democracy to Cleon’s “mob-oratory”. It was only natural, then, that aspiring politicians saw the advantages of proficiency in persuasive speaking. Fortunately for them, a new intellectual movement had arisen during this century, one which extended philosophical enquiry into fields for the most part untouched by the Presocratic philosophers, whose interests had largely lain in explaining the physical world. Amongst the topics now brought under examination were “moral, political, and social philosophy ... philosophy of language and epistemology” (McKirahan, 1994:354). One of the first of this new breed of intellectuals was the Sicilian rhetorician Gorgias, who wrote an *Encomium of Helen*, defending a woman whose actions were considered indefensible, to demonstrate that “λόγος δυνάστης μέγας ἐστίν, ὃς μικροτάτῳ σώματι καὶ ἀφανεστάτῳ θειότατα ἔργα ἀποτελεῖ” (8).²⁶ The whole point of the argument was to advertise his own rhetorical skills, and not surprisingly, his strategy worked. Rhetoric – the art of persuasion – was soon a requisite for any aspiring politician.

By the time of Cicero’s childhood, oratory²⁷ was firmly established as part of an upper class Roman education. It was taught as a subject at school, where boys learned its purpose, its three main kinds – “partim in iudiciis versari, partim in deliberationibus; esse etiam genus tertium, quod in laudandis aut vituperandis hominibus poneretur” (Cic. *De Orat.* 1.141)²⁸ – and the five divisions into which the skill and activities of the orator fell, namely “*inventio*, seeking out ideas or lines of argument; *collocatio*, structure and organisation; *elocutio*, diction and style; *actio*, physical delivery; and *memoria*, memory” (Everitt, 2003:28). After his sixteenth birthday Cicero was further tutored by some of the foremost speakers of the time – Quintus Mucius Scaevola the Augur, Crassus Orator and Antonius Orator, all three of whom feature as participants in Cicero’s dialogue *De*

²⁶ “Speech is a powerful master, who brings about the most divine deeds with the smallest and most imperceptible body.”

²⁷ The Latin translation of Greek ῥητορική. Quintilian prefers the Greek term to the Latin one, since he feels that *oratoria* is not an adequate or equivalent translation of the Greek (2.14.1-4).

²⁸ “One kind is found in law courts, another in deliberations; there is a third kind also, which is employed in praising or insulting people.”

Oratore. Nor did his rhetorical training end there, since a fledgling Roman politician would generally start his career in the law courts, building up a reputation in the forum before ever he entered the Senate.

Rhetoric was not, however, restricted to the political and legal domains. Quite naturally, it found its way even into literature. Although rhetoric initially meant nothing more than “the art of speaking” (*An Intermediate Greek-English Lexicon*, 1889: 717), it came to acquire other associations. In the ancient world the predominant view was that the purpose of oratory lies “aut in persuadendo aut in dicendo apte ad persuadendum” (Quint. *Instit. Orat.* 2.15.3),²⁹ although Aristotle, calling it a τέχνη “skill”, “knowledge”, “craft” (*Rhet.* 1.1354a), argues that “οὐ τὸ πείσαι ἔργον αὐτῆς, ἀλλὰ τὸ ἰδεῖν τὰ ὑπάρχοντα πιθανὰ περὶ ἑκάστων” (*Rhet.* 1.1355b).³⁰ In Cicero's *De Oratore* the character of Crassus claims that “primum oratoris officium esse, dicere ad persuadendum accommodate” (1.138).³¹ On the topic of whether oratory is an art, Crassus denies that there is an art of oratory, “si ars ita definitur ... ex rebus penitus perspectis planeque cognitis, atque ab opinionis arbitrio seiunctis, scientiaque comprehensis” (*De Orat.* 1.108).³² On the other hand, however, he concedes that since the tricks and techniques used by public speakers “animadversa ac notata, verbis designata, generibus illustrata, partibus distributa sunt” (*De Orat.* 1.109)³³ it is possible to call oratory an art of sorts, if not in the strictest, most scientific sense.³⁴ The two most important features of rhetoric, then, involve its being an art or a τέχνη and its ability to persuade.³⁵

Since the τέλος or purpose of rhetoric is “bene dicere” (Quint. *Inst. Orat.* 2.15.38),³⁶ it follows that rhetoric is concerned with the aesthetics of language. It does not deal with the arrangement of arguments and themes alone, but also with the arrangement of words.

²⁹ “Either in persuasion or in speech fit to persuade.”

³⁰ “Its work is not persuasion, but the knowledge of the possible means of persuasion in each case.”

³¹ “The first duty of an orator is to speak in a manner likely to persuade.”

³² “If art is so defined ... as comprised of matters that have been thoroughly studied and clearly understood, separate from the influence of conjecture, and within the domain of expert knowledge.”

³³ “Have been observed and noted, defined in words, made clear by categorisation, and divided into parts.”

³⁴ Cf. the *Concise Oxford Dictionary* (1964: 1070) definition of rhetoric as “the art of persuasive or impressive speaking or writing; language designed to persuade or impress”. *Chambers Dictionary of Etymology* (2002: 925) agrees, defining it as the “art of using language, especially to persuade or influence others”.

³⁵ *Vide* Quintilian 2.15 for a substantial discussion of the ancient definitions of rhetoric.

³⁶ “To speak well.”

Sit enim frequentissime aspera et dura et dissoluta et hians oratio si ad necessitatem ordinis sui verba redigantur, et ut quodque oritur ita proximis, etiam si vinciri non potest, alligetur. Differenda igitur quaedam et praesumenda, atque ut in structuris lapidum inpolitorum loco quo convenit quodque ponendum. Non enim recidere ea nec polire possumus quo coagmenta se magis iungant, sed utendum iis qualia sunt, eligendaeque sedes.

Quint. *Instit. Orat.* 8.6.62f.³⁷

In other words, literary texts will be examples of an order that is not natural, because they are intended to be aesthetically pleasing. Elegance is achieved through painstaking craftsmanship, where attention is paid to the placement of words so that they may each have their proper effect, that the sentence may flow euphoniously from one word to the next, and so that a natural rhythm is created:

neque alio <consilio in> ceris Platonis inventa sunt quattuor illa verba, quibus in illo pulcherrimo operum in Piraeum se descendisse significat, plurimis modis scripta <quam ut> quo ordine quodque maxime faceret experiretur.

Quint. *Instit. Orat.* 8.6.64³⁸

Ancient texts are accordingly particularly dangerous when it comes to studying Latin and Greek as they were used in everyday situations. The job of an author was to produce beautiful prose or poetry, and so the word order reflected in their works will be, to a greater or lesser degree, influenced by stylistic considerations.

This is not to say that ancient literary works are devoid of any value when it comes to studies of Latin syntax. In fact, in light of the communicative perspective of Latin grammar, rhetoric can be most useful in the understanding and explanation of unusual word orders. After all, Panhuis's approach stresses that Latin sentences are organised according to the message that the speaker wishes to convey, and rhetoric is all about effectively exploiting different word order patterns depending on the speaker's intention. However, these are not – as Quintilian points out above – the usual word order patterns, but deviations from the norm. If we want to establish basic word order in Latin, we must do so by turning to texts that are as close to the vernacular as possible.

³⁷ “For speech would very often be grating, harsh on the ear, lax and disjointed if words were restricted in accordance with the demands of their natural order, where as each arises so it is joined to the next, even if no connection can be made. So certain words must be put off or moved earlier, just as each block of rough-hewn stone in a building must be placed where it fits. For we can't cut or polish words when they have been placed side by side to make them fit better, but taking them as they are we must choose their places.”

³⁸ “Nor was there any other reason why, on Plato's wax tablets, those four words (in which – in his best work – he says that he went down to the Piraeus) were found written in many ways, than that he wished to test which order would be most effective.”

The *Rhetorica Ad Herennium*, an early first century B.C. work on oratory incorrectly attributed to Cicero, states that there are three main registers or styles, namely *figura gravis*, *figura mediocris* and *figura attenuata*. The first of these, *figura gravis*, is the grand style, “quae constat ex verborum gravium levi et ornata constructione” (*Rhet. Her.* 4.11),³⁹ whilst, as its name suggests, *figura mediocris* lies between the opulence of *figura gravis* and the simplicity of *figura attenuata*. This last style is defined as “id quod ad infumum et cottidanum sermonem demissus est” (*Rhet. Her.* 4.14)⁴⁰ – so it should be possible to find examples of straightforward and unembellished Latin prose.

The authors selected as paradigm examples of natural Latin are usually Plautus, Caesar and Cato the Elder, since, as Panhuis (1982:5) claims, they are not prone to “beltristic ornamentation”. Indeed, Cato should be a prime candidate for investigations of Latin word order, as Cicero points out that Cato, Pictor and Piso, as the early Roman historians, do not understand “quibus rebus ornetur oratio – modo enim huc ista sunt importata – et, dum intellegatur, quid dicant, unam dicendi laudem putant esse brevitatem” (*De Orat.* 2.53).⁴¹ However, Cato's works were still written to be published, and so his prose is bound to be more self-conscious than it would be in informal and relaxed speech.⁴² Panhuis (*op. cit.*, p. 5) motivates his choice of Plautus and Caesar by pointing out that Plautus “writes for a general popular audience in a language that, in spite of the verse-form sounds like everyday speech”, whilst likening Caesar's style to that of the early Roman annalists discussed in *De. Orat.* 2.53f.. Indeed, Cicero describes the style of Caesar's commentaries as “nudi ... recti et venusti, omni ornatu orationis tamquam veste detracta” (*Brutus* 262).⁴³ On the other hand, though, as a playwright Plautus writes in verse, which means that his sentences, and the order of the words within them, are automatically constrained to some degree by metrical requirements. Pinkster (1990:186) gives four examples of intentional violations of Latin word order – all of them come from poetry. This is not to say that Plautus produced such monstrosities as Ovid's “grandia per

³⁹ “Which consists of a polished and elaborate arrangement of dignified words.”

⁴⁰ “That which stoops to the lowest, everyday speech.”

⁴¹ “The ways in which speech is embellished – for those were only recently introduced here – and, as long as what they say is understood, they think the one virtue of narration is brevity.”

⁴² After all, Tacitus, writing in the lofty genre of history some three hundred years later, uses the incredibly grandiose periphrasis “virilis sexus stirpem” (*Ann.* 1.58.6) to say “baby boy”, and infamously avoids calling a spade a spade (*ibid.*, §65.7)

⁴³ “Plain ... unaffected and elegant, stripped bare of all oratorical embellishments.”

multos tenuantur flumina rivos” (*Rem.* 445),⁴⁴ but metre is metre and places demands upon language.

Caesar, too, is not as free from rhetoric as is often assumed to be the case. A brief look at an excerpt from one of his commentaries illustrates the point:

Erant hae difficultates belli gerendi, quas supra ostendimus, sed multa Caesarem tamen ad id bellum incitabant: iniuriae retentorum equitum Romanorum rebellio facta post deditionem, defectio datis obsidionibus, tot civitatum coniuratio, in primis, ne hac parte neglecta reliquae nationes sibi idem licere arbitrarentur. Itaque cum intellegent omnes fere Gallos novi rebus studere et ad bellum mobiliter celeriterque excitari, omnes autem homines natura libertati studere et condicionem servitutis odisse, priusquam plures civitates conspirarent, partiendum sibi ac latius distribuendum exercitum putavit.

BG 3.10⁴⁵

The passage is straightforward, the first sentence a catalogue of the many considerations that encouraged Caesar’s campaign against the Veneti. When this list is broken down into clausulae, the careful design that has gone into its making becomes apparent, with the causes – *rebellio*, *defectio* and *coniuratio* – placed either at the beginning of their clausula or the end, the two most significant positions in a Latin clause. *in primis* also stands out strongly, with a clausula to itself, highlighting the fact that Caesar has, of course, his priorities in order. In the next sentence, he follows the catalogue of causes with an εἰκός-type argument. First he appeals to his own firsthand experience of the Gallic temperament – undeniably warlike – and then makes the sweeping assertion that all people desire freedom and hate the state of slavery. As a result, it was only reasonable that he should dispose his army in such a way as to confine the uprising before it had a chance to spread. Note the dual alliteration of “*priusquam plures civitates conspirarent*”, echoed by “*partiendum*” and “*putavit*” a little later. Although at first glance the excerpt seems plain and forthright, the product of a well-organised mind, it has its fair share of “unobtrusive artifice” (Chandler, 2005:1), so subtly

⁴⁴ “Great rivers are dispersed through many streams.”

⁴⁵ “There were these difficulties of waging war, which we have shown above. However, many reasons, urged Caesar to battle: the unjust detention of Roman knights, the insurrection that came after surrender, desertion despite the fact that hostages had been given, the conspiracy of so many states, and above all, if this were ignored, the other nations would think that the same licence had been granted to them. So, since almost everyone knows that the Gauls have a passions for new things, being easily and swiftly stirred to war, and that all humans by nature desire liberty and hate the condition of servitude, he thought that the army should be divided up and stationed far and wide before more states joined in.”

done that the reader “does not receive the normal signals which would alert the ‘rhetorical immune system’” (*ibid.*).

If Caesar, praised for his “pura et inlustri brevitatem” (Cic. *Brutus* 262),⁴⁶ is to be viewed with suspicion because of his imperceptible manipulation of the reader, then Cicero may seem a far more unlikely source of colloquial Classical Latin. He was known as one of the leading advocates of his time, making his name through forensic oratory, and wrote several works on oratory, the largest one being his three-volume *De Oratore*. However, we are fortunate enough to have an extensive collection of Cicero's private correspondence – letters to friends and family. These will almost certainly *not* have been written with an eye for epideixis. In particular the letters to his brother Quintus and to his good friend Titus Pomponius Atticus are quite intimate and casual as the result of his close relationship with them. Consequently, they are probably the closest surviving examples of colloquial Classical Latin available to us.

Covering the years between 68 and 44 B.C., the *Epistulae Ad Atticum* are personal, informal, chatty, their subject matter ranging from Cicero's inflamed eyes to the birth of his son Marcus, from his row with his brother and nephew to the political machinations of the *Triumviri*. The letters are confidential, meant for Atticus's eyes only – Cicero describes one as “epistula quam nolo aliis legi” (1.16.8)⁴⁷ – and were “written without any idea of future publication” (Shackleton-Bailey, 1999:1). Admittedly, he is aware that letters may miscarry, and at one point establishes a simple code for sensitive political correspondence (2.19.5, 20.5), but that does not change the fact that these are private letters intended solely for Atticus. As a result, there is no audience who needs to be impressed by the author's verbal virtuosity – after all, he is writing to someone “cum quo a condiscipulatu vivebat coniunctissime” (Nepos *Att* 5.3).⁴⁸ Indeed, Atticus is not merely an old friend but is assured by Cicero that “a me fraterne amari” (1.5.8),⁴⁹ and is later described as “in publica re socius et in privatis omnibus conscius et omnium meorum

⁴⁶ “Simple and clear brevity.”

⁴⁷ “A letter which I don't want anybody else to read.”

⁴⁸ “With whom he had lived very happily from boyhood.”

⁴⁹ “Loved by me like a brother.”

sermonum et consiliorum particeps” (1.18.1).⁵⁰ According to Nepos (*Att.* 16.2) not even his brother Quintus was dearer or closer to Cicero.

It is Atticus to whom Cicero turns when exiled and on the verge of a nervous collapse, and Atticus whose advice he seeks when agonising over the right decision to make during the civil war. In these letters we see Cicero taking great emotional and mental strain, especially those at the beginning of Book III, written during his departure from Rome on his way into exile. Short and to the point, they reflect the pressure Cicero is under: “utinam illum diem videam cum tibi agam gratias quod me vivere coëgisti! adhuc quidem valde me paenitet” (3.3.1),⁵¹ “plura scribere non possum; ita sum animo perculso et abiecto” (3.2.1),⁵² “me, mi Pomponi, valde paenitet vivere” (3.4.1),⁵³ “ego vivo miserrimus et maximo dolore conficior” (3.5.1),⁵⁴ “hoc adfirmo, neminem umquam tanta calamitate esse adfectum, nemini mortem magis optandam fuisse” (3.7.2).⁵⁵

However, letters remain letters: they are written, and give Cicero enough time to think about what he says and how he says it. Even though he claims that his distress has robbed him of “omnis partis mentis” (3.7.3),⁵⁶ he still manages to include an example of polyptoton with “neminem”, “nemini” earlier in the same letter. A master orator does not forget the tricks of his trade so easily, especially one who has written a work on oratory in three books, and scaled the *cursus honorum* through a reputation made as a public speaker in the forum.

Nevertheless, the letters are definitely not examples of elegant and polished rhetoric, nor are they written in a high style. Cicero uses the adjective *bellus* “good” fairly frequently, e.g. “homines belli” (1.1.2), “valde bella est” (4.6.4), “bellum ἀκροτελεύτιον” (5.21.3), as well as the adverb *belle* “well”, e.g. “posita ita belle est” (1.1.5), “fecisse non belle” (5.17.6), “plane belle se habere” (12.37.1). This is the

⁵⁰ “An ally in public affairs, a confidant in all private matters, and a participant in all my speeches and councils.”

⁵¹ “I wish I might see the day when I will thank you for making me live! Right now I certainly regret it.”

⁵² “I can’t write any more, I am so beaten down and depressed.”

⁵³ “It pains me greatly to live, dear Pomponius.”

⁵⁴ “My life is miserable and I’m worn out by profound grief.”

⁵⁵ “I maintain this – that nobody ever was affected by so tremendous a disaster, and that nobody should have begged more for death.”

⁵⁶ “All my faculties.”

vernacular form of Classical Latin *bonus*,⁵⁷ which it ultimately ousted in Vulgar Latin and the Romance languages, e.g. French *beau*, *bel*, *belle*, Italian *bello*, *bella*. In narrative pieces describing senatorial circuses or mayhem in the streets he will not even use full sentences, e.g.

itaque a. d. III Id. Nov. cum Sacra via descenderem, insecutus est me cum suis.
clamor, lapides, fustes, gladii; et haec improvisa omnia. discessi in vestibulum
Tetti Damionis.

4.3.3, italics added⁵⁸

This is even more casual than the example of *figura attenuata* in the *Rhetorica Ad Herennium*. So it would seem that Cicero's letters to Atticus are as close to colloquial Classical Latin as it is possible to get, and so can be used to work out the basic word orders relatively free from rhetorical interference.

⁵⁷ *bellus* is actually derived from the diminutive form of *bonus*, just as *sol* was ousted by its own diminutive *soliculus*, a process "extremely common in late Latin" (Grandgent, 1907:10).

⁵⁸ "So, on 11 November, as I was coming down the Sacra Via, he and his gang came after me. Uproar, flying stones, cudgels, even swords! And all this with no warning whatsoever! I slipped into Tettius Damio's forecourt."

CHAPTER 2

GoldVarb

Whilst linguistic variation often appears to be a messy welter of a variety of options, sociolinguistic studies have demonstrated that this is not actually the case and that variation is the result of “structured heterogeneity” (Weinreich et al., 1968, quoted by Bayley, 2002:117), i.e. it is systematic and governed by contextual factors.¹ These factors can range from the linguistic, e.g. phonetic environment, tense and case, to the non-linguistic, e.g. social class, gender and ethnicity. However, the choice of one variant over another is often not determined by one factor alone, but by a combination of factors.² These concepts are fundamental to the quantitative approach to variation analysis in linguistics.

Tagliamonte (2006) provides a brief history of the development and use of variable rule analysis in sociolinguistics, which first arose “as a quantitative extension of generative phonological analysis and notation” (p. 130, citing Labov, 1969, 1972). The expression “variable *rule*” is perhaps somewhat misleading, since “variable rules are actually ‘the probabilistic modelling and the statistical treatment of discrete choices and their conditioning’” (p. 131, quoting Sankoff, 1988). The programmes most commonly used to analyse linguistic variation are VARBRUL for Windows and its Macintosh counterpart GoldVarb, since they are “deliberately designed” to deal with the problems of language data. Standard statistical models, such as analysis of variance, or ANOVA, rely upon an equal distribution of tokens of the variable amongst the factors (Bayley, *op. cit.*, p. 124, Tagliamonte, *op. cit.*, p. 137), but when it comes to language, this is rarely – if ever – the case. Take, for example, the table below, excerpted from my own data on Cicero’s Latin:

¹ Young and Bayley’s “principle of quantitative modeling” (1996, quoted by Bayley, *op. cit.*, p. 117). This principle essentially refers to the “likelihood of co-occurrence of a variable form and any one of the contextual features in which we are interested” (Bayley, *op. cit.*, p. 118).

² The “principle of multiple causes” (Young and Bayley, 1996, quoted by Bayley, *op. cit.*, p. 118), which states that “it is unlikely that any single contextual factor can explain the variability observed in natural language data” (Bayley, *op. cit.*, p. 118).

Word Order in Cicero's *Letters to Atticus*

(1)

Dependent variable adjective/noun order in Cicero's Latin [1 = adjective follows; 0 = adjective precedes] with independent variable whether the noun phrase is part of a preposition phrase or not [n = not part of preposition phrase; e = part of preposition phrase]

Group		1	0	Total	%
2 (3)		1	0		
n	N	229	373	602	77.6
	%	38.0	62.0		
e	N	47	127	174	22.4
	%	27.0	73.0		
Total	N	276	500	776	
	%	35.6	64.4		
TOTAL	N	276	500	776	
	%	35.6	64.4		

In this example, it is clear that the tokens are not evenly distributed across the factors n and e. Some 602 of the 776 examples of noun phrases are *not* part of a preposition phrase, as opposed to only 174 that are. It is this sort of difficulty that GoldVarb has been designed to overcome.

Another inconvenient feature of language is the fact that interaction or “dependence among factor groups” is a “natural aspect of speech” (Tagliamonte, *op. cit.*, p. 137). An extreme example is given below:

(2)

Horizontally - type of adjective is: possessive [p], descriptive [c], demonstrative [d], reflexive ipse [r]

Vertically - language of the adjective is: Latin [L], Greek [G]

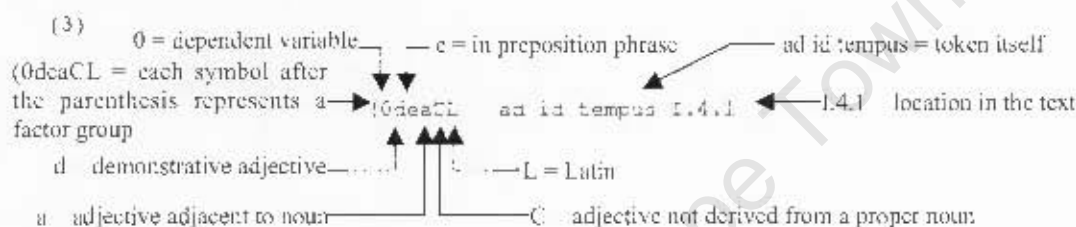
· CROSS TABULATION · 09/23/07 20:40:59

	p		c		d		r		Σ	%
	+	-	+	-	+	-	+	-		
L 0:	97	45:	261	69:	130	82:	10	59	498	65
1:	118	55:	115	31:	28	18:	7	41	268	35
Σ:	215	:	376	:	158	:	17	:	766	
G 0:	0	--:	2	20:	0	--:	0	--	2	20
1:	0	--:	8	80:	0	--:	0	--	8	80
Σ:	0	:	10	:	0	:	0	:	10	
Σ 0:	97	45:	263	68:	130	82:	10	59	500	64
1:	118	55:	123	32:	28	18:	7	41	276	36
Σ:	215	:	386	:	158	:	17	:	776	

From here, it can be seen that the only Greek adjectives Cicero uses are descriptive, hence the factor group language is dependent upon the factor group type. As a result, the

two groups need to be recoded, either by omitting the language group, or by collapsing it into the type group, and splitting the descriptive adjectives between Latin descriptives and Greek descriptives. GoldVarb is able to reveal such interactions through its cross-tabulation function, although Bayley notes that the programme “does not provide a convenient way to test for interactions among factor groups” (*op. cit.*, p. 130).

How does GoldVarb run? First, it is necessary to note all the tokens of the variable under investigation, specifying the contextual factors for each token. So, for example, “ad id tempus” (1.4.1) would be coded as



Once the token file has been created, and all the data coded, GoldVarb then works out the weights of each factor relative to the other factors within its group, i.e. it calculates the likelihood of the variable occurring in conjunction with that particular factor (Bayley, *op. cit.*, p. 126).³ The factor weights range between 0.00 and 1.00, with values of 0.5 and up favouring the application value, the variant that is chosen as the application of the variable rule (Bayley, *op. cit.*, p. 126 f., Tagliamonte, *op. cit.*, p. 253). GoldVarb also calculates the log likelihood of the analysis, a measurement of how well it fits the data. The closer the log likelihood is to 0, the better the correlation. It further works out the input value, which Tagliamonte defines as “an overall measure of rule application” (*op. cit.*, p. 264). So, returning to example (1), the distributional analysis shows that the dependent variable 1 – the adjective follows its noun – occurs 35.6% of the time, whilst in example (4) below, which is GoldVarb’s stepping-up run for the same construction,

³ Bayley (*op. cit.*, p. 126) gives the formula used:

$$p = \frac{p_0 \times \dots \times p_n}{[p_0 \times \dots \times p_n] + [(1 - p_0) \times \dots \times (1 - p_n)]}$$

where p is the factor weight or probability, p_0 is the input probability and p_1 to p_n are “the applicable factor weights from each of the factors in the model”. Tagliamonte defines the input as the “overall tendency of rule application” (*op. cit.*, p. 156) – how likely is it that the variable rule will apply regardless of the contextual factors?

with 1 set as the application value, the input hovers around 0.343, which means that the overall likelihood that Latin adjectives will follow their nouns is roughly 0.34, a figure which closely corresponds with the distributional analysis.

Finally, the programme tests the significance of each factor group, to see how much influence each has on the application of the variable rule investigated. It does so by means of a step-up/step-down analysis, which ultimately supplies “statistical significance, relative strength and constraint ranking of factors” (Tagliamonte, *op. cit.*, p. 140). During the stepping-up run, the programme determines which factor group is most important, the group that “makes the most significant change to the model when it is added or subtracted from the rest” (*ibid.*).

```
(4)
Adjective/noun order with three factor groups, step-up
Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:
Convergence at Iteration 2
Input 0.356
Log likelihood = -505.088

----- Level # 1 -----

Run # 2, 5 cells:
Convergence at Iteration 5
Input 0.344
Group # 1, Type of adjective: possessive [p]; Latin descriptive [L],
demonstrative [d], reflexive [r], Greek descriptive [G]
-- p: 0.699, L: 0.457, d: 0.291, r: 0.572, G: 0.884
Log likelihood = -469.848 Significance = 0.000

Run # 3, 2 cells:
Convergence at Iteration 4
Input 0.354
Group # 2, Whether noun phrase is in a preposition phrase or not: not part of
a preposition phrase [n], part of a preposition phrase [e]
-- n: 0.528, e: 0.403
Log likelihood = -501.391 Significance = 0.008

Run # 4, 3 cells:
Convergence at Iteration 4
Input 0.354
Group # 3, Adjacency of adjective to noun: adjacent [a], separated by another
modifier within the same phrase [s], separated by an element not part of the
same phrase [h]
-- a: 0.529, s: 0.428, h: 0.374
Log likelihood = -501.395 Significance = 0.027

Add Group # 1 with factors pLdrG
```

A Multivariate Approach

```
----- Level # 2 -----  
  
Run # 5, 9 cells:  
Convergence at Iteration 5  
Input 0.342  
Group # 1 -- p: 0.704, L: 0.453, d: 0.291, r: 0.590, G: 0.871  
Group # 2 -- n: 0.532, e: 0.390  
Log likelihood = -465.451 Significance = 0.005  
  
Run # 6, 14 cells:  
Convergence at Iteration 5  
Input 0.343  
Group # 1 -- p: 0.694, L: 0.460, d: 0.290, r: 0.567, G: 0.881  
Group # 3 -- a: 0.521, s: 0.458, h: 0.383  
Log likelihood = -467.787 Significance = 0.135  
  
Add Group # 2 with factors ne  
  
----- Level # 3 -----  
  
Run # 7, 24 cells:  
Convergence at Iteration 5  
Input 0.341  
Group # 1 -- p: 0.700, L: 0.457, d: 0.290, r: 0.585, G: 0.868  
Group # 2 -- n: 0.533, e: 0.387  
Group # 3 -- a: 0.522, s: 0.457, h: 0.375  
Log likelihood = -463.133 Significance = 0.099  
  
No remaining groups significant  
  
Groups selected while stepping up: 1 2  
Best stepping up run: #5  
-----
```

With each level of the step-up analysis more factor groups are tested in combination with each other. So at Level #0, none of the groups are tested, whilst at Level #1, each group is tested individually. At the end of Level #1, Group #1 (type of adjective) is added as the most significant factor group, since its log likelihood was the lowest, and so the group fits best with the data. In Level #2, Group #1 is retained and the remaining two groups tested in conjunction with it. The combination of Groups #1 and #2 turns out to have a better log likelihood than the combination of Groups #1 and #3, so Group #2 (preposition phrases) is added as the second-most important factor. At Level #3, however, when all three factor groups are run together, the significance level does not improve, and so Group #3 (adjacency) is not selected. Once the step-up has concluded, the run that corresponds best to the data is Run #5, the Level #2 run which combined both Groups #1 and #2. The selected factor groups are given at the end of the step-up in order of significance (in this example, it also happens to be numerical, but that is not always the case).

The step-down analysis is the reverse of the step-up. Instead of starting out with no factor groups, it begins by establishing the fit of the model when all factor groups are present, and then rejects the group “whose loss least significantly reduces the likelihood” (Tagliamonte, *op. cit.*, p. 143).

(5)

Adjective/noun order with three factor groups, step-down
Stepping down...

----- Level # 3 -----

Run # 8, 24 cells:
Convergence at Iteration 5
Input 0.341
Group # 1 -- p: 0.700, L: 0.457, d: 0.290, r: 0.585, G: 0.868
Group # 2 -- n: 0.533, e: 0.387
Group # 3 -- a: 0.522, s: 0.457, h: 0.375
Log likelihood = -463.133

----- Level # 2 -----

Run # 9, 6 cells:
Convergence at Iteration 5
Input 0.352
Group # 2 -- n: 0.530, e: 0.398
Group # 3 -- a: 0.530, s: 0.425, h: 0.363
Log likelihood = -497.260 Significance = 0.000

Run # 10, 14 cells:
Convergence at Iteration 5
Input 0.343
Group # 1 -- p: 0.694, L: 0.460, d: 0.290, r: 0.567, G: 0.881
Group # 3 -- a: 0.521, s: 0.458, h: 0.383
Log likelihood = -467.787 Significance = 0.004

Run # 11, 9 cells:
Convergence at Iteration 5
Input 0.342
Group # 1 -- p: 0.704, L: 0.453, d: 0.291, r: 0.590, G: 0.871
Group # 2 -- n: 0.532, e: 0.390
Log likelihood = -465.451 Significance = 0.099

Cut Group # 3 with factors ash

----- Level # 1 -----

Run # 12, 2 cells:
Convergence at Iteration 4
Input 0.354
Group # 2 -- n: 0.528, e: 0.403
Log likelihood = -501.391 Significance = 0.000

Run # 13, 5 cells:
Convergence at Iteration 5
Input 0.344
Group # 1 -- p: 0.699, L: 0.457, d: 0.291, r: 0.572, G: 0.884
Log likelihood = -469.848 Significance = 0.005

All remaining groups significant

Groups eliminated while stepping down: 3

A Multivariate Approach

Best stepping up run: #5
Best stepping down run: #11

So at Level #2, when Group #1 is cut, the log likelihood drops from -463.133 to -497.260 – definitely not an improvement. On the other hand, the best log likelihood at this level occurs when Group #3 is rejected, confirming it as the factor group that plays the least significant rôle in determining the ordering of nouns and adjectives in Cicero's Latin. At Level #1, the log likelihood worsens when either Groups #1 or #2 are cut – not the dramatic jump when Group #1 is discarded. As a result, the programme concludes that both remaining groups are significant, and isolates the best stepping down run, to be compared with the best stepping up run, Run #11. This should be identical with Run #5, which it is, bar the significances.

It should be mentioned at this point that even though GoldVarb has been designed to cope with unevenly distributed linguistic data, there are still distributions that it cannot handle, namely the "KnockOuts". These are factors which are categorical, i.e. they only ever use one variant of the variable. For example, in my Honours dissertation, interrogative Latin adjectives invariably preceded the noun they modified, and were thus classified as knockout factors by the programme. GoldVarb cannot conduct its variable analysis in this case "because the data, so configured, is not variable" (Tagliamonte, *op. cit.*, p. 152).

However, the programme remains a useful tool for the analysis of linguistic variation, and not just within the field of sociolinguistics. Since Latin word order appears so free on the surface, GoldVarb seems a natural choice for an in-depth statistical analysis that will determine the relative importance of the factors affecting the constructions under investigation. That is, after all, what the programme was intended to do.

CHAPTER 3

Methodology

3.1 Methodology

Out of the sixteen books of Cicero's letters to Atticus, I analysed the syntax within only the first four. Admittedly, some of the tokens for the genitive construction were drawn from Book 9, since I re-used the data from my Honours dissertation. Each of the letters in these first four books was then broken down into *commata*, or *incisa*. "Incisum ... erit sensus non expleto numero conclusus" (Quint. *Instit. Orat.* 9.4.122), i.e. it is a semantically discrete but metrically incomplete unit. This was done on the grounds that the comma can perhaps be considered the unit of composition of a Latin sentence: the flexibility of the syntax encourages sentences to be constructed from word groupings based on sense. Moreover, *commata* tend not to be merely semantically coherent, but also syntactically complete. For example

Kal. Iuniis
eunti mihi Antium,
et gladiatores M. Metelli cupide relinquenti,
venit obviam tuus puer.

Cic. *Att.* 2.1.1¹

The first comma here gives the date on which the action of the sentence happened, using an ablative of time, whilst the second and third *commata* are participle phrases agreeing with the indirect object *mihi*. The second comma explains that Cicero was *en route* to Antium, the third provides more information regarding the reasons for his journey. The last comma is the main clause – the meeting with one of Atticus's slaves carrying a letter for Cicero. Sense and syntax coincide.

Once the letters had been broken down into their constituent *commata*, I created token files for each of the four constructions whose word order patterns were under investigation. These were the order (1) of adjective and noun, (2) of genitive and noun, (3) of adverb and verb, and (4) of object and verb. I then noted each realisation of these constructions as well as a number of potential factors, clustered into factor groups, that

¹ "On the Kalends of June your boy met me as I was going to Antium, wanting to leave M. Metellus's gladiators behind."

might affect the ordering of the constituents. In all cases, if the modifier preceded the noun it scored 0, and if it followed it scored 1.

3.2 Constructions

3.2.1 Adjectives

The tokens for the adjective construction were drawn from Book 1 of *Letters to Atticus*. The additional information considered in conjunction with the adjectives was comprised of the following five factor groups:

1. the type of adjective – possessive (p), e.g. *tuum puerum* (1.1.1), *deliciae nostrae* (1.5.5), demonstrative (d), e.g. *illa omnia* (1.4.2), *eandem rationem* (1.11.1), descriptive (c), e.g. *di immortales* (1.16.1), *bonus homo* (1.19.10), or the emphasising adjective *ipse* (r), e.g. *ipso tempore* (1.1.1);
2. whether the NP formed part of a preposition phrase (e), e.g. *cum magna pecunia* (1.1.2) or not (n), e.g. *barbatuli iuvenes* (1.14.5);
3. the proximity of the two constituents to each other – they were either adjacent (a), e.g. *omni virtute* (1.5.1), or separated by some other modifier within the same phrase (s), e.g. *haec praepropera pensatio* (1.1.1), or separated by an element that did not form part of the NP (h), e.g. *bonis utimur tribunis pl.* (1.14.6);
4. whether the adjective was derived from a proper noun (P), e.g. *operae Clodianae* (1.14.5), *Autronianam domum* (1.13.6) or not (C), e.g. *di boni* (1.16.5), *amor vester* (1.17.2);
5. and finally the language – Latin (L) or Greek (G) – in which the adjective occurred, as Cicero dropped in several Greek adjectives, e.g. *ornamenta γυμνασιώδη* (1.6.2).

Interrogative adjectives were omitted, since in my Honours dissertation they were invariably placed before the noun, and were thus knockout factors (McLachlan, *op. cit.*, p. 22).

After my first run of the data, the results for the best stepping up and down runs were identical, except for the significance, which increased from 0.005 in the stepping-up run

to 0.099 in the stepping-down run. I then cross-tabulated my data to spot any factor interaction, and modified my factor groups in the following ways in an attempt to lower the significance level for the stepping-down run:

1. I collapsed Group 5 (language) into Group 1 (type of adjective), since Cicero used Greek for descriptive adjectives alone, which changed the factors within Group 1 to possessive (p), demonstrative (d), the emphasising adjective *ipse* (r), and Latin descriptive adjectives (L) and Greek descriptive adjectives (G);
2. I cut Group 4 (common or proper), since it was not considered a significant group. Groups 2 and 3 remained unchanged. This left me with just 3 factor groups.

3.2.2 Genitives

The factor groups for the genitive construction were:

1. the type of constituent used in the genitive, namely a noun (n), e.g. *discipulorum arculas* (2.1.1), a demonstrative pronoun (d), e.g. *istorum tempora* (2.7.3), a possessive pronoun, e.g. *eius actiones* (9.15.2), or a gerund, e.g. *spem fruendi* (9.13.6);
2. as above, whether the NP formed part of a preposition phrase (e), e.g. *ex eius nuntio* (9.17.1), or not (n), e.g. *pacis fides* (Att. 9.19.3);
3. again as above, the proximity of the two constituents to each other – they were either adjacent (a), e.g. *adulescentulorum studiis* (2.1.3), or separated by some other modifier within the same phrase (s), e.g. *istorum inter ipsos dissensio* (2.7.3), or separated by some an element that did not form part of the NP (h), e.g. *expectationem dedisti convivii* (2.12.2);
4. whether the noun in the genitive case was a common (C), e.g. *mentis errore* (3.13.2), or a proper noun (P), e.g. *L. Reguli libertus* (Att. 3.17.1);
5. and last of all, the function of the genitive – was it possessive (g), e.g. *animus eius* (9.10.6), partitive (p), e.g. *plena laudis* (2.3.3), subjective (s), e.g. *reditu Caesaris* (9.10.10), objective (o), e.g. *cives amantes patriae* (9.19.3), appositional (a), e.g. *magnitudo animi* (2.15.2), or a genitive of quality (q) *res sane magni consilii* (2.3.3)?

These tokens were taken from Books 2 and 3. As stated above, I also re-used the tokens taken from Book 9 that I had used in my Honours dissertation.

In a subsequent run, attempting to improve the significance of the best stepping-up and down runs, Group #4 was collapsed into Group #1, since the distinction between common and proper affected only nouns, not pronouns, nor gerunds. Appositional genitives and genitives of quality in Group #5 were also collapsed into one factor q due to the similarity of their functions. However, the run succeeded only in worsening the significances, so it was discarded.

3.2.3 Adverbs

For the ordering of adverbs and verbs within VPs the factors were:

1. the part of speech the adverb was derived from – was it derived from a noun (n), e.g. *magno opere confirmo* (2.6.1), a preposition or preposition phrase (p), e.g. *venit obviam* (2.1.1), an adjective (a), e.g. *plane cessasse* (2.7.1), a verb (v), e.g. *statim des* (2.5.3), or was it a plain adverb (d) such as *nunc audi* (2.9.4);
2. again as above, the proximity of the two constituents to each other – they were either adjacent (a), e.g. *prorsus abhorret* (2.6.1), or separated by some other modifier within the same phrase (s), e.g. *vix misereque sustento* (2.6.1), or separated by some element that was not part of the same phrase (h), e.g. *vero haec addita cura vix mihi vitam reliquam facit* (3.8.2);
3. again as above, the language of the adverb, either Latin (L) or Greek (G), e.g. *ἐπὶ σκολῆς scribe* (2.5.3) ;
4. and lastly, the function of the adverb – was it an adverb of place (p), e.g. *proficisci istinc* (2.6.2), manner (m), e.g. *subcontumeliose tractatur* (2.7.3), or time (t), e.g. *statim consequare* (3.1.1)?

All tokens of this particular construction come from Books 2 and 3.

3.2.4 Objects

Finally, object/verb word order had the following six factor groups:

1. the weight of the object – light (l), e.g. *litteras darem* (4.1.1), or heavy (h), e.g. *tibi, praesertim iam prope praesenti, scriberem* (4.4.1);

2. the communicative dynamism of the object – was it rhematic (r), e.g. *mihi mittas de tuis librariolis duos aliquos* (4.4a.1), or thematic (t), e.g. *mihi mittas de tuis librariolis duos aliquos* (4.4a.1);
3. the proximity of the two constituents to each other – they were either adjacent (a), e.g. *gratias agebat* (4.9.1), or separated by some other modifier within the same phrase (s), e.g. *habet Pomptinam, Velinam, Maeciam* (4.15.9), or separated by some element that was not part of the same phrase (h), e.g. *mihi sermo desit* (4.11.2);
4. the case of the object – accusative (a), e.g. *copias nosti* (4.16.6), genitive (g), e.g. *misereri mei* (4.5.2), dative (d), e.g. *scribas mihi* (4.8.2), or ablative (b), e.g. *consiliis utitur* (4.3.2);
5. the degree of animacy of the object – low (l), e.g. *offendes designationem* (4.4a.1), mid (m), e.g. *libros illustrarunt* (4.8.2), or high (h), e.g. *vide absolutum Gabinium* (4.19.1);
6. and finally the thematic rôle of the object – was it a theme (T), e.g. *voramus litteras* (4.11.2), a patient (P), e.g. *me principem nominavit* (4.1.7), an experiencer (E), e.g. *Caesari displicere* (4.17.3), a recipient or beneficiary/maleficiary (F), e.g. *tibi mitto* (4.11.2), an instrument (I), e.g. *lepore suo detineat* (4.15.2), a possessum (p), e.g. *habet Pomptinam* (4.15.9), a source (s), e.g. *Dyrrachio sum profectus* (4.1.4), or a goal (G), e.g. *Brundisium veni* (4.1.4). Since source and goal had few tokens, and source was also a knock out factor, only ever occurring before the verb, they were collapsed together into one group (L) for the actual run.

The data for each construction was then fed into GoldVarb, which performed the requisite statistical analysis. The results, which are included in an appendix, are discussed below.

CHAPTER 4

Discussion of results

4.1 Adjectives

The straightforward distributional analysis of the 776 adjective tokens found in Book 1 of the *Epistulae Ad Atticum* shows that the most common order is AdjN, found exactly 500 times (64.4%), even though it is disharmonic for OV languages.¹ The uneven spread of tokens within each factor group also clearly reveals the natural “‘lumps’ and ‘clumps’, ‘hollows’ and ‘dips’, in the data” (Tagliamonte, 2006:139). So, within Group #1, the descriptive adjectives alone make up nearly 50% of the data (386 occurrences), whilst the reflexive adjective *ipse* comprises a scant 2.2% (17 tokens).

According to the distributional analysis, the following orders dominate in terms of frequency of occurrence:

1. the type of adjective – Np, LN, dN (very marked), rN and NG (very marked);
2. AdjN both when the NP is part of a PP, and when it is not; however, the preference for AdjN order rises by eleven percent when part of a PP;
3. the proximity of the two constituents to each other – again, AdjN throughout, with the dominance of the order increasing from 61.9% when adjacent to 71% when separated by an element that is also part of the same phrase, to 75.4% when separated by a constituent external to the NP.

Of these, the first two groups alone were selected as significant factors, with type of adjective considered more important than subordination of the NP within a PP. This was the only construction for which proximity was not significant.

However, the cross tabulation of Groups #1 and #3 shows that in all except three cases, the separated modifiers demonstrated a substantially increased preference for AdjN order than the adjacent modifiers. Only the reflexives which had undergone hyperbaton and the Greek descriptive adjectives which were separated but remained in the same phrase broke from this pattern, both invariably following their head noun. This may have been because

¹ AdjN order is marked for Devine and Stephens; however, it is also the underlying order in which the construction is generated before movement in the functional theory (*op. cit.*, p. 484; see above).

of the insufficient number of tokens available for these cells – one for the reflexives and two for the Greek descriptive adjectives. Moreover, in the first book of the *Epistulae Ad Atticum*, the Greek adjectives never underwent hyperbaton. Again, the cross tabulation of Groups #2 and #3 shows the same increase in dominance of the AdjN order whenever the adjectives were separated from the noun, regardless of whether they were separated by an element from the same NP or not. Further, a glance at the log likelihoods of the best stepping-up and stepping-down runs (-465.451 for both), and the runs in which all three factor groups were retained (-463.133, again for both) shows that the likelihood actually improved when proximity was not cut, although the significances worsened. Since the likelihood did not show a dramatic improvement, though, and since the likelihoods of all other runs containing it were poorer than those without it, proximity was discarded.

Turning to the statistical analysis now, the input of 0.342 (application value 1) in the best stepping-up and down runs tallies well with the distributional analysis's figure of 35.6% for occurrences of NAdj order, confirming that the dominant order for Cicero is indeed AdjN. As for the factor weights, in Group #1 the orders Np (very marked), LN, dN (very marked) and NG (very marked) all correspond to the orders predicted from their frequency of occurrence. The only exception to this is Nr, with a factor weight of 0.590, despite the fact that in the distributional analysis rN is favoured 58.8% of the time.² However, it should be borne in mind that the actual tokens are the result of a combination of different factors from different factor groups, whilst GoldVarb's statistical analysis considers the influence of each factor in isolation. In other words, the statistics are indicative of the importance of the factor itself, and may not correspond to the examples provided, where other factors come into play.

Pinkster (1990:185) provides a useful table in which he gives the positions of attributes, or modifiers, relative to their head nouns, the data being drawn from generalisations made in Latin grammars.

² This also reverses the order in my Honours dissertation, where the reflexive scored 0.490, and so preferred rN, in accordance with its distribution. The factor weight was, however, very close to 0.5, so rN and Nr were virtually in free variation with each other.

A Multivariate Approach

Type of Attribute	Position of Attributes with regard to Heads	
	Before	After
Demonstrative pronoun	+	
Possessive pronoun		+
“Determining” adjective		+
“Qualifying” adjective	+	
Complex Attribute		+
Numeral	+	
Noun in the genitive		+

Pinkster, 1990:185

So it seems that the rules for adjective location depend on whether or not they are determining or qualifying, i.e. whether they limit (determine) the reference of their head noun or merely describe it. Descriptive adjectives should accordingly precede the noun – and the Latin descriptive adjectives certainly do, with a weight of 0.457. They also score 0 some 69.4% of the time in the frequency distribution, although the Greek adjectives show a marked preference for NAdj order. The possessives and demonstratives, on the other hand, should follow their noun, as the possessives indeed do. The demonstratives, however, have a factor weight of 0.290, and so definitely precede the noun they modify. However, according to the table, the demonstrative pronoun tends to occur in premodifier position, so perhaps it is unsurprising that the adjective also comes before its noun.

The distinction between determining and qualifying adjectives may lend itself to Devine and Stephens’s pragmatic analysis of adjective movement in terms of topic and focus (*op. cit.*, p. 483 f). Because the information that qualifying adjectives carry is descriptive, it may often be new and significant, an important property of the noun which helps to set it apart from other nouns with the same reference. For example, in the phrase “*praepropera pensatio*” (Cic. *Att.* 1.1.1) the noun *pensatio* represents old information. More noteworthy in the context is the fact that Galba’s canvassing is premature. Since the qualifying adjective is more significant than the noun itself it occupies the focus position of FocNP, from which it c-commands the noun in the base NP. Similarly, demonstrative adjectives may be contrastive – *this* dog rather than *that* dog – or they may presuppose earlier information and act as topic links, and be moved all the way to the highest projection TopNP, whilst the noun moves to FocNP.

On the other hand possessive adjectives may be contrastive – *my* cat not *your* cat – but more frequently, the noun they describe is important, not to whom it belongs. For example, in “avunculus tuus” (Cic. *Att.* 1.1.3) there is no question that Caecilius is anybody’s uncle but Atticus’s, or again in “commodo tuo” (1.1.4), where the emphasis is not on at whose leisure Atticus’s return is made, but on the fact that it is to be made when he pleases, not when it is a nuisance. Consequently, the noun is scrambled into TopNP where it can c-command the adjective in FocNP (Devine and Stephens, *op. cit.*, p. 483).

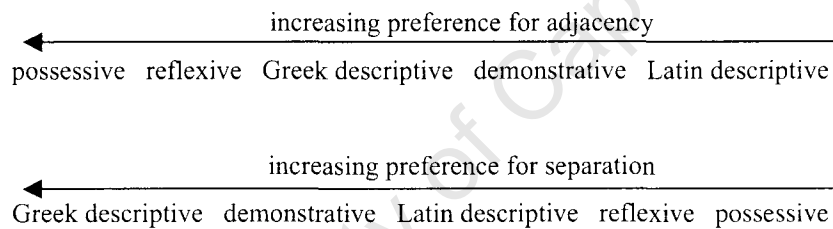
A tentative rule for adjective stacking may also be worked out from a comparison of the distributional analysis with the cross tabulation of Group #1 with Group #3. The cross tabulation shows that when adjacent, the orders Np, LN, dN, rN and NG are favoured, when separated, pN, LN, dN, rN and NG dominate, and when the noun and adjective undergo hyperbaton, pN, LN, dN and Nr prevail. From the total number of tokens for each type of adjective given in the cross tabulation, the frequency with which they occur in each of the six different possible positions can be calculated. For example, there are 215 tokens of the possessive adjective. 182 of these are found adjacent to the noun they modify, 20 are separated by an element in the same phrase and 13 by an external element. These figures can be further broken down into adjacent pN (77 tokens, or 35.81% of the total of 215), adjacent Np (105, or 48.84%), separated pN (12, or 5.85%), separated Np (8, or 3.72%), hyperbaton pN (8, or 3.72%) and hyperbaton Np (5, or 2.33%). The frequency percentages for all the Group #1 factors are given in the table below:

	possessive	Latin descriptive	demonstrative	reflexive	Greek descriptive
adjacent Σ	84.65%	71.54%	74.68%	82.35%	80%
adj. premodifier	35.81%	48.94%	60.13%	47.06%	20%
adj. postmodifier	48.84%	22.61%	14.56%	35.29%	60%
separated Σ	9.30%	18.62%	18.99%	11.76%	20%
sep. premodifier	5.58%	12.5%	17.09%	11.76%	N/A
sep. postmodifier	3.72%	6.12%	1.90%	N/A	20%
hyperbaton Σ	6.05%	9.84%	6.33%	5.88%	N/A
hyp. premodifier	3.72%	7.98%	5.06%	N/A	N/A
hyp. postmodifier	2.33%	1.86%	1.27%	5.88%	N/A

Comparing the total percentages for each position, all of the adjectives are more likely to be found immediately adjacent to their noun than in any other position, with

possessives showing the greatest preference for juxtaposition (84.65%) and Latin descriptive adjectives the least (71.54%). Greek adjectives are separated by an NP-internal element 20% of the time, but possessives only 9.3% of the time. And Latin descriptive adjectives undergo hyperbaton 9.84% of the time, which contrasts with Greek descriptive adjectives, which never undergo hyperbaton. However, the hyperbaton order is least relevant for adjective stacking, since stacking involves adjacency to and separation from the noun by one or more modifiers within the same NP, not by a constituent external to the NP.

Once hyperbaton has been discarded, looking only at the adjacent and separated constructions, the adjective types can be ranked in order of preference for each position as follows:



Possessives and reflexives both tend to stay as close to the noun as possible, with the highest preferences for adjacency and the lowest for separation. In both cases, their totals are separated by little more than 2%, which might pose a potential problem for determining which type of adjective comes where when stacked. However, a glance at the dominant orders for adjacency solves the difficulty – reflexives are far more likely to precede when adjacent, and possessives to follow. This gives the order rNp.

Greek adjectives, although they are also found adjacent to their noun a good proportion of the time, are the most likely of all the adjectives to be separated from their noun. At all times, they show a marked preference for postmodifier position, and categorically follow the noun when separated. The order of adjectives can now be elaborated: rNpG. As regards the Latin descriptives and the demonstratives, the Latin adjectives are less likely to be immediately adjacent to the noun, but the demonstratives are more likely to be separated – though only by 0.37%. This gives a tentative final stacking order dLrNpG,

which is in accordance with Greenberg's Universal 20, which states that when demonstrative and descriptive adjectives are premodifiers of the noun, they will occur in the order demonstrative-descriptive-noun. A frequency count of stacked adjectives in Book I of the letters to Atticus showed that this stacking rule held up 36% of the time; however, it was also contradicted just as often, with orders such as the disharmonic LdN, e.g. *novam quandam potentiam* (1.19.4) – this unsurprising, given the fact that the demonstratives are more likely to be adjacent to the noun, and that they are only 0.37% more likely than the Latin descriptives to be separated from the noun.

Returning to the statistical analysis of the construction, according to the factor weights for Group #2, when the NP is not part of a PP, NAdj is favoured, although the score is only 0.532, which means that AdjN is almost as likely to occur. But when the NP is part of a PP then the factor weight drops to 0.390, showing a strong preference for p + AdjN. This confirms my findings in my Honours dissertation, and again stands in opposition to the distribution of both plain adjective-noun order and adjective-noun order in preposition phrases cited by Elerick (in Herman, *op. cit.*, p. 69):

NAdj 64%	AdjN 36%	prep + NAdj 79%	prep + AdjN 21%
			Elerick, <i>ibid.</i>
NAdj 38%	AdjN 62%	prep + NAdj 27%	prep + AdjN 73%
			Cic. <i>Att.</i> 1

As I pointed out previously (McLachlan, *op. cit.*, p. 24), Elerick's source here is Livy, who would have been in his mid teens at the time of Cicero's death in 43 B.C., whilst the first book of his history of Rome is usually dated to roughly 27 B.C. on the basis of internal evidence. There is always the possibility of language change, a suggestion that Elerick himself puts forward (*ibid.*, p. 70) when explaining some differences between Caesar and Livy. However, it should be noted that Livy's order is actually somewhat more conservative than Cicero's, despite Livy being the later author. Devine and Stephens (*op. cit.*, p. 405) observe that "the rules for adjective location underwent a diachronic change" at this time, shifting from Cato's rigorous NAdj order in *De Agricultura* to freer variation between pre- and postmodifier position in later Latin. Cicero's preference for AdjN in his letters may have to do with this change. It may also be worth noting that Livy's *Ab Urbe Condita* is a published history, a formal work of an

elevated genre,³ and so prone to linguistic conservatism, whilst the letters to Atticus are the casual correspondence of very good friends. As a result, they ought to reflect the more progressive norms of colloquial speech which the literary works do not.

Whatever the reason, the distributional analysis and the input value both favour the disharmonic order AdjN, although the factor weight for NPs that do not form part of a PP tends to the harmonic NAdj. On the other hand, the weight for NPs that are part of a PP demonstrates a marked preference for the disharmonic prep + AdjN, which means that Elerick's principle of Harmonic Phenotypic Linearization does not hold for Cicero's letters. However, in his study of adjective-noun order when the noun forms the genitive element of a genitive construction, Elerick states that the dominant word order is AdjN (in Herman, *op. cit.*, p. 70). Significantly, his source here is Cicero himself, suggesting that the AdjN pattern was a feature of his Latin and a result of "the strong variation one observes between individual authors and text types" (Pinkster, 1990:188).

4.2 Genitives

For the genitive construction, a total of 470 tokens were drawn from Books II, III and IX of Cicero's *Letters to Atticus*, the fewest number of tokens for any of the four constructions studied, though taken from more books. Unlike the adjectives, for which the dominant sequence was clearly AdjN, the most common order for the genitives was NGen (cf. Pinkster's table above; also the default order for Devine and Stephens, *op. cit.*, p. 381), but this in only 251 of the cases – a mere 53.4% of the time. The near free variation between NGen and GenN corresponds with Elerick's findings (in Herman, *op. cit.*, pp. 69 ff.), which he once again puts down to language change from an original GenN in "prehistoric Latin" (*ibid.*, p. 71) to NGen in Romance, perhaps originally motivated by considerations of semantics or of weight, with the HPL adding impetus to the shift (prep + NGen is harmonic).

The dominant orders suggested by the distributional analysis are given below:

³ See Ch. 1, n. 42 on p. 22 above.

1. the type of genitive – Nn (but quite free), pN (very marked) and dN; as for the genitive gerunds, neither vN nor Nv were preferred, since both occurred exactly three times;
2. NGen when the NP was not part of a PP, and the disharmonic prep + GenN, but again both orders are in relatively free variation with their inverses;
3. the proximity of the two constituents to each other – NGen throughout, with the preference for this order increasing from 50.1%⁴ when adjacent to 60.5% when separated by an element that was also part of the same phrase, to 67.7% when separated by a constituent external to the NP;
4. whether genitive nouns were common or proper – proper nouns favoured GenN, common nouns NGen; when collapsed into Group #1 in the second run, these orders remained unchanged;
5. and lastly, the function of the genitive – all functions favoured NGen, bar subjective genitives, for which neither order was dominant, and the possessive genitives, which preferred NGen, but only by one more token than GenN; partitive genitives were the most strongly biased in favour of NGen (73.5%).

The three groups selected as significant after GoldVarb had performed the statistical analysis were, in order of importance, Group #1, type of constituent in the genitive, Group #4, common or proper noun in the genitive, and Group #3, adjacency. Surprisingly, Group #2, subordination within a PP, was cut, despite the fact that it had been selected for the adjective construction, and that its two factors had displayed inverse orders in the genitive frequency distribution above. Further, the cross tabulations of Group #2 with each of the other groups shows that there is a definite trend for the proportion of GenN occurrences to increase in PPs – once again the HPL does not hold for Cicero. Admittedly, the runs which included this factor group had the worst log likelihood values in Levels #2, #3 and #4 of the stepping-up run, nor were their significances good. It is possible that there may simply have been insufficient examples of genitive constructions within PPs (82 tokens) for the group to be considered significant.

⁴ Admittedly, this figure hardly suggests that NGen is favoured substantially over GenN.

The fact that Group #5, function, was discarded is somewhat unexpected, since one might well expect that the syntactic rôle played by a genitive would determine its location relative to its head. On the other hand, Devine and Stephens (*op. cit.*, p. 314) note that in their study of Latin word order there were

no immediately obvious rules for the order of nominal arguments and adjuncts relative to the head, at least no rules that are expressible in terms of the familiar syntactic categories. For instance, it is not the case that objective genitives always [sic] follow the noun and subjective genitives always precede the noun; or that possessive genitives always precede the noun and partitive genitives always follow the noun; or that deverbal nouns always take a genitive to the left and relational nouns always take a genitive to the right. In fact it is typically the case that within each category of genitive we investigated we found that different words had different distributions ... [We can't] say that word order in noun phrases is random. The data analysis tells us explicitly that it isn't random: it varies according to lexical item.

It would therefore seem that the syntactic function of the genitive is quite unimportant, whilst lexical restrictions are significant. However, coding for individual words would have led to an incredibly large and unwieldy token file for GoldVarb to handle, with more than its share of knockout factors and poorly distributed data, so I did not include lexical item as a factor group.

The statistical analysis confirms that NGen is the preferred order, with an input value of 0.535 (application value 1) for the best stepping-up and down runs. The factor weights confirm the predicted trends in the distributional analysis, as well as establishing that when the genitive is a gerund the order GenN is favoured slightly, with a score of 0.464. Interestingly, the possessive pronoun *eius* is heavily weighted to precede its possessum (0.256), which conflicts with Pinkster's table of attributes (*op. cit.*, p. 185; see above), where it is said to follow its head. Whilst the difference might be due to the fact that the table is derived from frequency of occurrence, rather than on a statistical analysis, my distributional analysis corresponds closely to the factor weight, with the possessive pronoun preceding the noun in 21 out of 30 tokens. The table is admittedly based upon generalisations made by "the grammars" (Pinkster, *op. cit.*, p. 185), which presumably drew upon a wide range of authors in order to come up with their generalisations – perhaps Cicero's word order is once again idiosyncratic.

To return to the statistical analysis, the only other area in which it differs from the distributional analysis is the order given for adjacent genitives – GenN (0.470), not NGen (which occurred 50.1% of the time) – whilst the distributional trend of increased preference for NGen with increased distance from the noun is borne out. Separated genitives score 0.557, whilst those undergoing hyperbaton have a factor weight of 0.635. This is in direct opposition to the pattern seen with the adjectives, in which the further apart the adjective and the noun were, the more strongly the premodifier position was favoured. In fact, this is the only construction in which an increase in distance leads to a greater likelihood of the modifier following its head – both adverbs and objects tend to precede the verb when separated from it (see below).

However, the basic orders for adjectives, adverbs and objects are AdjN, AdvV and OV respectively, whilst for genitives it is NGen. It becomes immediately apparent that separation, and hyperbaton in particular, reinforces the basic order of the constituents. The simplest explanation for this is that when the words are immediately adjacent to each other, their grammatical relationship is plain to see and easily understood. However, when the words are separated, the link between them is harder to discern, especially when undergoing full hyperbaton. It makes sense, then, that in this case the neutral order will be preferred to an unusual one, since it is predictable and will thus be easier to parse. Quintilian himself observes that for hyperbaton, “[f]elicissimus tamen sermo est cui et rectus ordo ... contigit” (*Instit. Orat.* 9.4.27).⁵ It is, after all, an unusual rhetorical device, “the most distinctively alien feature of Latin word order” (Devine and Stephens, *op. cit.*, p. 524), and as such calls into play the principle of Preemptive Markedness, where “one instance of markedness is constrained by the other” (Elerick, in Calboli, *op. cit.*, p. 561). Of course, given the characteristic variability of Latin, the rule does not always hold, but the pattern is definitely there, even in the case of the adjectives, the only construction for which degree of adjacency was not considered a significant factor.

⁵ “The best style is the one in which the normal order ... is found.”

4.3 Adverbs

Adverbs were very strongly weighted for AdvV in the distributional analysis, with 501 out of 605 tokens preceding the verb. Indeed, not one of the thirteen factors favours VAdv, though they all have a small percentage of tokens where the adverb does follow its verb. So the predicted order throughout is AdvV.

At first glance, it may seem futile trying to search for variation within a construction with such an overwhelming preference for one order over the other. However, the cross tabulations show that all is not as uniform as it appears on the surface. Admittedly, there are no cells which favour VAdv, but there are four in which both orders are in free variation. These are:

1. Group #1 and Group #2 – separated adverbs which are either PPs or derived from prepositions;
2. Group #1 and Group #4 – adjective-derived adverbs of place;
3. Group #2 and Group #3 – Greek adverbs that have undergone hyperbaton;
4. and Group #2 and Group #3 once again – Greek adverbs that are separated from their verb by another element within the same VP.

It should be observed, though, that in all four cases, the number of tokens available is minimal. However, interestingly enough, the factor weights calculated by GoldVarb show that preposition-derived adverbs and separated adverbs both tend to follow their verbs, as they score 0.653 and 0.598 respectively.

Only two of the four factor groups are considered significant by the statistical analysis (application value 1). Once again the most important group is related to the type of modifier, in this case, the part of speech that the adverb was derived from. The factor weights for this group demonstrate that not all adverbs precede their verb, since noun, preposition and adjective-derived adverbs have weights of 0.537 and up, although the verb-derived adverbs and the unclassified adverbs such as *nunc* “now” hold true to the AdvV order predicted by the distributional analysis, particularly the unclassified adverbs, with a score of 0.208.

While the adjective-derived adverbs are close enough to 0.5 for VAdv to be co-dominant with AdvV, the preposition-derived adverbs are firmly weighted for postmodifier position. This may seem anomalous in light of the input value of 0.143, as well as the fact that in the distributional analysis they precede the verb 74.3% of the time. However, these adverbs are not just derived from prepositions, e.g. *praeterea*, *coram*, but are very often full-blown preposition phrases, e.g. *in propinquis locis* (2.1.4), *per Macedoniam* (3.8.1). These phrases are undeniably heavier than most verbs, and so undergo heavy constituent shift, moving to the right of the lighter verbs. Pinkster (1990:167), however, argues against internal complexity or weight of phrases triggering this shift. Taking relative clauses as an example, he points out that they are more likely to precede the verb, despite being complex and therefore heavy, than they are to follow. However, his argument is based once more on a straightforward distributional analysis. See also the discussion of the results for object-verb order below.

The second significant factor group selected by GoldVarb was the degree of adjacency. This time, though, separated adverbs did not reinforce the neutral order AdvV, nor did they improve upon the score of the adjacent adverbs. In fact, separated adverbs are slightly more likely to follow the verb than are adjacent adverbs. This contradicts the claim made above, that the greater the distance between the head and its adjunct, the stronger the tendency will be to the neutral order. On the other hand, adverbs that undergo hyperbaton are firmly slanted towards AdvV order, with a factor weight of 0.336, which is in accordance with the claim. It is worth noting that a mere 9.4% of the 605 adverb tokens were separated, whilst 55.4% of them were adjacent and 35.2% had undergone hyperbaton. It is possible, then, that the factor weight for separated adverbs might drop with an increase in the number of tokens of this particular construction.

Yet again, the syntactic function of the modifier – whether it was an adverb of manner, place or time – turned out to be unimportant. However, a brief glance at its factor weights in those runs where it was not cut suggests that adverbs of place consistently favour postmodifier position, whilst adverbs of manner favour premodifier position. The cross-tabulation of Groups #1 and #4, providing the composition of each function, i.e. which types of adverb make up the bulk of each of the factors place, manner and time, show that 82 of the 145 adverbs of place are preposition-derived. Preposition-derived adverbs, as

discussed above, are the most likely of all the types of adverbs to follow the verb, due to their weight, and since – unsurprisingly – they make up just over half of the adverbs of place, it is only natural that these adverbs should follow their verbs.

This seems to conflict with Devine and Stephens' observation that adverbials of place "are well attested in preverbal position" (*op. cit.*, p. 69); however, the frequency distribution for adverbs of place in Cicero's letters shows that 106 or 73.1% of them do occur before the verb. On the other hand, Devine and Stephens also noted that adjuncts of place might be generated either VP-externally or VP-internally depending on whether they were one of three types, namely event-external, event-internal or framing (*op. cit.*, p. 92). It is possible that those adverbs of place that follow their verbs are event-internal adverbs locating just a part of the event, rather than the event as a whole, which have not been raised, but left instead in tail position.

As for the adverbs of manner, the cross-tabulation of Groups #1 and #4 shows that 184, or roughly 68%, of them are adjective-derived – another type that tends to follow the verb. But adverbs of manner have factor weights ranging between 0.427 at the lowest and 0.451 at the highest, indicating that they favour the order AdvV. However, with a score of 0.537, the neutral AdvV is almost as likely to occur as the favoured VAdv, which leaves the adverbs of manner with sufficient freedom to tend to the neutral order. Also, adverbs of manner tend to be just one word, rather than an entire phrase, and a good number of them are relatively light, e.g. *plane*, *late*, *sane*, and so not heavy enough to trigger heavy constituent shift and be moved to the end of the phrase, inverting the preferred AdvV order. Furthermore, when cross-tabulated with the other significant factor group, #2, just over a third of the adverbs of manner undergo hyperbaton, the factor weighted most heavily for AdvV, which presumably also helps to bring their score down.

4.4 Objects

For this last construction, the neutral order, based on the frequency distribution, was OV, which occurred 78.9% of the time in a total of 531 tokens. The input value of 0.167

(application value 1) showed an even stronger preference for an OV order, and out of the twenty-one factors only one, in the distributional analysis, favoured VO (objects in the genitive case).

According to the distributional analysis, the following orders dominate in terms of frequency of occurrence:

1. both light and heavy objects precede the verb, although the OV order is substantially more marked for light objects (81.5%) than it is for heavy (60%);
2. OV again for both rhematic and thematic objects, with rhematic objects on the one hand preceding 66.1% of the time, whilst the thematic objects preceded the verb 90.1% of the time;
3. the proximity of the two constituents to each other – again, OV throughout, with the dominance of the order increasing from 70% when adjacent to 85.3% when separated by an element that was also part of the same phrase, to 89.9% when undergoing hyperbaton;
4. OV for objects in the accusative, dative and ablative cases, with dative the most marked (84%) and ablative the least (69.2%), but VO for objects in the genitive case, an order which occurred 57.1% of the time;
5. OV throughout animacy of the object, with highly animate objects preceding the verb in 82.7% of their cases, low animacy objects 79.5% and mid animacy objects 63.9%;
6. and last of all, OV throughout once more for thematic rôle.

Of these six factor groups, three were selected as significant. In order of importance these were Group #2, communicative dynamism of the object, Group #3, adjacency, and Group #1, weight. Case, animacy and thematic rôle were all discarded. The cross-tabulations of Groups #4 and #6 and Groups #5 and #6 highlight the extreme interaction of these three groups. In the first cross-tabulation, it becomes apparent that Latin case is used to some degree as a means of marking thematic rôle. So, for example, the rôle of instrument is marked by dative and ablative, and is never found in either the accusative or the genitive, whilst the vast majority of patients and themes are in the accusative. Animacy also interacts with thematic rôles, and thus, by extension, with case. There are

naturally no highly animate themes, instruments or sources and goals, nor any patients and experiencers with low animacy.

With so much factor interaction, it is hardly surprising that GoldVarb discarded these groups, with the least significant being case, then thematic rôle, and finally animacy. Thematic rôle also corresponds to the semantic function of the object, and so, like all other function-based factor groups, it is cut. The interplay between case and function so clearly demonstrated in the cross-tabulation here may perhaps explain the rejection of function as a significant factor. Function in Latin is encoded in the case endings, after all, and so does not need to be expressed by the organisation of the constituents, unlike English with its scant inflectional morphology and correspondingly rigid word order.

The choice of communicative dynamism of the object as the most important factor determining its position relative to the verb supports Panhuis's theory that Latin word order is based upon the communicative purpose of the speaker. The factor weights also confirm that thematic elements do indeed come earlier in the sentence, since thematic objects score 0.350, whilst rhematic objects are more likely to follow the verb, with a factor value of 0.670. Quintilian too agrees that the focus position⁶ in a Latin sentence is at the end:

Saepe tamen est vehemens aliquis sensus in verbo, quod si in media parte sententiae latet, transire intentionem et obscurari circumiacentibus solet, in clausula positum adsignatur auditori et infigitur ...

Instit. Orat. 9.4.29⁷

In short if the object contains old information, then the sentence will end with a verb, since "in verbis enim sermonis vis est" (Quint. *Instit. Orat.* 9.4.26), but if the object contains new and important information, then the order of constituents is more likely to be VO, which is marked thanks to its break with the strongly preferred OV order. Further, clause-final position is more salient than clause-medial position, witness the fact that the metre in clausulae is worked out from the end of each clausula. Thus, the new and

⁶ Cf. the tree structures of Devine and Stephens (*op. cit.*), which have TopXP the highest, and so leftmost, projection of any construction, with FocXP the projection that has immediate scope over XP (cf. pp. 12 f. above). This ensures that any element carrying focus will be as close to the end of the construction as possible.

⁷ "Often there is some strong significance in a word, which, if it lies concealed in the middle of the sentence, usually loses its meaning and is overshadowed by the surrounding words, but when placed in a clausula it is marked out to be heard and remembered ..."

important information contained in the rhematic object is highlighted when it is shifted to the end of the sentence.

Adjacency is once again found to play an important rôle in the arrangement of constituents. For the third time, the weighting for the neutral word order improves with increasing distance between the object and the verb, dropping from 0.649 (and so VO) for adjacent objects to 0.384 for separated objects to 0.326 for objects undergoing hyperbaton. As for object weight, light objects favour OV, but only just, with a score of 0.464, whilst heavy ones are definitely more likely to follow the verb, scoring 0.740. This conflicts with Pinkster's assertion (1990:167; see above) that weight has less to do with word order than whether the constituent is an argument or satellite – that is, essential or inessential to the understanding of the sentence – does. Instead, it ties in with the heavy constituent shift, which moves heavy elements to the ends of their clauses. Further, in the cross-tabulation of Groups #1 and #2, it can be seen that roughly two-thirds of the heavy objects are rhematic, and so more likely to tend to VO order. Considering that heavy objects contain more information than light objects, the interaction between communicative dynamism and weight is hardly surprising, providing a further motivation for the movement of heavy objects to a post-verbal position.

Conclusion

Although Latin word order has long been held to be free, recent studies have shown that it is governed by both syntactic and communicational considerations, the former largely on the phrasal level, the latter more on the sentence level. A variety of approaches has been taken to investigating word order patterns, ranging from the typological, e.g. Elerick's principle of Harmonic Phenotypic Linearization, to the syntactical, e.g. the same scholar's study of gapping and Preemptive Markedness in Latin, to the communicational, e.g. de Jong's article on Topic and Focus factors underlying subject placement, and Panhuis's Communicative Perspective in Latin sentences. Devine and Stephens's (2006) book exhibits a mix of approaches, combining empirical studies of individual words with theoretical discussions that seek to explain the patterns observed in their data from syntactic, semantic and pragmatic points of view.

This dissertation too has been based upon empirical evidence and statistical analysis, but there are several differences between it and other studies. Firstly, I believe that if we want to describe word order in Classical Latin, we need to start by establishing neutral or basic patterns. This is best accomplished by turning to informal Latin, of which Cicero's private correspondence is probably the best surviving example. Most of the other surviving texts from the Classical period are literary works and so far more conscious of their style than the letters, especially given the ancient preoccupation with rhetoric and the resultant aesthetically pleasing cadences and linkages into which sentences could be shaped. The oft-cited Caesar, although his commentaries are simple reading, is hardly free from artifice – indeed, the very accessibility of his texts is what makes him a dangerous author to work with, since his manipulation of language (and the reader) is so skilled that it is virtually imperceptible. Plautus on the other hand is constrained to some degree by the metrical requirements of his plays, even though his Latin is definitely colloquial and informal. So, whereas many other studies draw upon literary works, I have concentrated solely upon four books of informal Classical Latin, in accordance with the sociolinguistic principle that the target of all linguistic enquiry is the vernacular, or the most relaxed and least self-conscious style.

A second caution is that a number of studies also take their information from a range of authors from different time periods. Panhuis uses Plautus (c. 254 B.C. – 184 B.C.) and Caesar (100 B.C. – 44 B.C.), Elerick uses Caesar, Cicero (106 B.C. – 43 B.C.) and Livy (59 B.C. – 17 A.D.), Devine and Stephens use a wide variety of authors, at one point comparing Cato the Elder (234 B.C. – 149 B.C.) with Columella (c. 4 B.C. – A.D. 70). Devine and Stephens and Elerick do acknowledge the possibility of language change – but that is precisely the point. Language is constantly changing and developing, and even in a literate society, where writing slows the pace of change, the differences in vocabulary and style between a text written two hundred years ago and a text written now are plainly visible. The practice of making observations based upon such a wide chronological spread of authors is surely more than a little linguistically unsound.

Further, most articles and books on Latin word order are content with deducing word order patterns from frequency distributions. However, by using the statistical linguistics programme GoldVarb, tailored to cope with the irregularities of language data, I have attempted to determine significant factors underlying the organisation of constituents within Latin sentences. From my results, it would appear that the arrangement of arguments, modifiers and adjuncts in relation to their heads is not free, but based upon a medley of syntactic and communicational constraints. At phrase level in particular the syntactic considerations appear to be strongest, with the type of modifier, i.e. the part of speech it either is, or is derived from, consistently selected as the most important factor. The syntactic function of modifiers is generally discarded, except in the case of adjectives, where function and type coincide. As for objects, grammatical function does not need to be marked by word order, since it is already contained in the inflectional morphology.

Indeed, in verb phrases, the arrangement of objects and verbs has very little to do with syntax and more to do with the communicational purpose of the sentence. Important information tends to be found in a clause-final position; thus, when an object has a high degree of communicative dynamism, it will gravitate towards a VO order, marked by the fact that it is the focus position of a Latin sentence, as well as by prosodic patterns and by the fact that it inverts the favoured order OV. Constituent weight is also important, with

Word Order in Cicero's *Letters to Atticus*

----- Level # 2 -----

Run # 5, 9 cells:

Convergence at Iteration 5

Input 0.342

Group # 1 -- p: 0.704, L: 0.453, d: 0.291, r: 0.590, G: 0.871

Group # 2 -- n: 0.532, e: 0.390

Log likelihood = -465.451 Significance = 0.005

Run # 6, 14 cells:

Convergence at Iteration 5

Input 0.343

Group # 1 -- p: 0.694, L: 0.460, d: 0.290, r: 0.567, G: 0.881

Group # 3 -- a: 0.521, s: 0.458, h: 0.383

Log likelihood = -467.787 Significance = 0.135

Add Group # 2 with factors ne

----- Level # 3 -----

Run # 7, 24 cells:

Convergence at Iteration 5

Input 0.341

Group # 1 -- p: 0.700, L: 0.457, d: 0.290, r: 0.585, G: 0.868

Group # 2 -- n: 0.533, e: 0.387

Group # 3 -- a: 0.522, s: 0.457, h: 0.375

Log likelihood = -463.133 Significance = 0.099

No remaining groups significant

Groups selected while stepping up: 1 2

Best stepping up run: #5

Stepping down...

----- Level # 3 -----

Run # 8, 24 cells:

Convergence at Iteration 5

Input 0.341

Group # 1 -- p: 0.700, L: 0.457, d: 0.290, r: 0.585, G: 0.868

Group # 2 -- n: 0.533, e: 0.387

Group # 3 -- a: 0.522, s: 0.457, h: 0.375

Log likelihood = -463.133

----- Level # 2 -----

Run # 9, 6 cells:

Convergence at Iteration 5

Input 0.352

Group # 2 -- n: 0.530, e: 0.398

Group # 3 -- a: 0.530, s: 0.425, h: 0.363

Log likelihood = -497.260 Significance = 0.000

A Multivariate Approach

Run # 10, 14 cells:
Convergence at Iteration 5
Input 0.343
Group # 1 -- p: 0.694, L: 0.460, d: 0.290, r: 0.567, G: 0.881
Group # 3 -- a: 0.521, s: 0.458, h: 0.383
Log likelihood = -467.787 Significance = 0.004

Run # 11, 9 cells:
Convergence at Iteration 5
Input 0.342
Group # 1 -- p: 0.704, L: 0.453, d: 0.291, r: 0.590, G: 0.871
Group # 2 -- n: 0.532, e: 0.390
Log likelihood = -465.451 Significance = 0.099

Cut Group # 3 with factors ash

----- Level # 1 -----

Run # 12, 2 cells:
Convergence at Iteration 4
Input 0.354
Group # 2 -- n: 0.528, e: 0.403
Log likelihood = -501.391 Significance = 0.000

Run # 13, 5 cells:
Convergence at Iteration 5
Input 0.344
Group # 1 -- p: 0.699, L: 0.457, d: 0.291, r: 0.572, G: 0.884
Log likelihood = -469.848 Significance = 0.005

All remaining groups significant

Groups eliminated while stepping down: 3
Best stepping up run: #5
Best stepping down run: #11

2. Genitives

· CELL CREATION · 08/29/07 21:21:25
.....
Name of token file: Genitive_II_III_IX.tkn
Name of condition file: Genitive_II_III_IX.cnd
(
; Identity recode: All groups included as is.
(1)
(2)
(3)
(4)
(5)
(6)
)

Number of cells: 54
Application value(s): 10
Total no. of factors: 17

Word Order in Cicero's *Letters to Atticus*

Group		1	0	Total	%
<hr/>					
1 (2)		1	0		
n	N	225	173	398	84.7
	%	56.5	43.5		
p	N	9	21	30	6.4
	%	30.0	70.0		
v	N	3	3	6	1.3
	%	50.0	50.0		
d	N	14	22	36	7.7
	%	38.9	61.1		
Total	N	251	219	470	
	%	53.4	46.6		
<hr/>					
2 (3)		1	0		
n	N	216	172	388	82.6
	%	55.7	44.3		
e	N	35	47	82	17.4
	%	42.7	57.3		
Total	N	251	219	470	
	%	53.4	46.6		
<hr/>					
3 (4)		1	0		
a	N	184	183	367	78.1
	%	50.1	49.9		
h	N	44	21	65	13.8
	%	67.7	32.3		
s	N	23	15	38	8.1
	%	60.5	39.5		
Total	N	251	219	470	
	%	53.4	46.6		
<hr/>					
4 (5)		1	0		
P	N	52	60	112	23.8
	%	46.4	53.6		
C	N	199	159	358	76.2
	%	55.6	44.4		
Total	N	251	219	470	
	%	53.4	46.6		
<hr/>					
5 (6)		1	0		
g	N	116	115	231	49.1
	%	50.2	49.8		
o	N	53	41	94	20.0
	%	56.4	43.6		

A Multivariate Approach

a	N	11	9	20	4.3
	%	55.0	45.0		
s	N	43	43	86	18.3
	%	50.0	50.0		
p	N	25	9	34	7.2
	%	73.5	26.5		
q	N	3	2	5	1.1
	%	60.0	40.0		
Total	N	251	219	470	
	%	53.4	46.6		

TOTAL	N	251	219	470	
	%	53.4	46.6		

Name of new cell file: .cel

· BINOMIAL VARBRUL · 08/29/07 21:21:34

Name of cell file: .cel

Averaging by weighting factors.
 Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:
 Convergence at Iteration 2
 Input 0.534
 Log likelihood = -324.689

----- Level # 1 -----

Run # 2, 4 cells:
 Convergence at Iteration 5
 Input 0.533
 Group # 1 -- n: 0.532, p: 0.273, v: 0.467, d: 0.358
 Log likelihood = -319.008 Significance = 0.010

Run # 3, 2 cells:
 Convergence at Iteration 4
 Input 0.534
 Group # 2 -- n: 0.523, e: 0.394
 Log likelihood = -322.398 Significance = 0.036

Run # 4, 3 cells:
 Convergence at Iteration 4
 Input 0.535
 Group # 3 -- a: 0.466, h: 0.645, s: 0.571
 Log likelihood = -320.771 Significance = 0.020

Word Order in Cicero's *Letters to Atticus*

Run # 5, 2 cells:

Convergence at Iteration 4

Input 0.534

Group # 4 -- P: 0.431, C: 0.522

Log likelihood = -323.254 Significance = 0.093

Run # 6, 6 cells:

Convergence at Iteration 5

Input 0.536

Group # 5 -- g: 0.467, o: 0.529, a: 0.515, s: 0.465, p: 0.706, q: 0.565

Log likelihood = -320.890 Significance = 0.184

Add Group # 1 with factors npvd

----- Level # 2 -----

Run # 7, 8 cells:

Convergence at Iteration 5

Input 0.533

Group # 1 -- n: 0.532, p: 0.270, v: 0.465, d: 0.359

Group # 2 -- n: 0.523, e: 0.392

Log likelihood = -316.694 Significance = 0.035

Run # 8, 9 cells:

Convergence at Iteration 5

Input 0.535

Group # 1 -- n: 0.531, p: 0.281, v: 0.497, d: 0.355

Group # 3 -- a: 0.468, h: 0.643, s: 0.556

Log likelihood = -315.491 Significance = 0.033

Run # 9, 5 cells:

Convergence at Iteration 6

Input 0.534

Group # 1 -- n: 0.538, p: 0.247, v: 0.433, d: 0.327

Group # 4 -- P: 0.394, C: 0.534

Log likelihood = -315.788 Significance = 0.012

Run # 10, 14 cells:

Convergence at Iteration 5

Input 0.535

Group # 1 -- n: 0.530, p: 0.279, v: 0.476, d: 0.371

Group # 5 -- g: 0.491, o: 0.503, a: 0.488, s: 0.447, p: 0.682, q: 0.536

Log likelihood = -316.365 Significance = 0.391

Add Group # 4 with factors PC

----- Level # 3 -----

Run # 11, 10 cells:

Convergence at Iteration 6

Input 0.534

Group # 1 -- n: 0.538, p: 0.248, v: 0.435, d: 0.331

Group # 2 -- n: 0.520, e: 0.408

Group # 4 -- P: 0.405, C: 0.530

Log likelihood = -314.202 Significance = 0.079

A Multivariate Approach

Run # 12, 12 cells:

Convergence at Iteration 6

Input 0.535

Group # 1 -- n: 0.537, p: 0.256, v: 0.464, d: 0.327

Group # 3 -- a: 0.470, h: 0.635, s: 0.557

Group # 4 -- P: 0.400, C: 0.532

Log likelihood = -312.683 Significance = 0.046

Run # 13, 17 cells:

Convergence at Iteration 8

Input 0.535

Group # 1 -- n: 0.540, p: 0.235, v: 0.463, d: 0.323

Group # 4 -- P: 0.393, C: 0.534

Group # 5 -- g: 0.515, o: 0.465, a: 0.446, s: 0.452, p: 0.643, q: 0.493

Log likelihood = -313.749 Significance = 0.540

Add Group # 3 with factors ahs

----- Level # 4 -----

Run # 14, 21 cells:

Convergence at Iteration 6

Input 0.535

Group # 1 -- n: 0.537, p: 0.256, v: 0.463, d: 0.330

Group # 2 -- n: 0.516, e: 0.425

Group # 3 -- a: 0.472, h: 0.625, s: 0.550

Group # 4 -- P: 0.408, C: 0.529

Log likelihood = -311.647 Significance = 0.160

Run # 15, 37 cells:

Convergence at Iteration 8

Input 0.536

Group # 1 -- n: 0.538, p: 0.244, v: 0.490, d: 0.321

Group # 3 -- a: 0.470, h: 0.627, s: 0.563

Group # 4 -- P: 0.397, C: 0.533

Group # 5 -- g: 0.515, o: 0.468, a: 0.446, s: 0.460, p: 0.629, q: 0.425

Log likelihood = -311.013 Significance = 0.650

No remaining groups significant

Groups selected while stepping up: 1 4 3

Best stepping up run: #12

Stepping down...

----- Level # 5 -----

Run # 16, 54 cells:

Convergence at Iteration 8

Input 0.536

Group # 1 -- n: 0.538, p: 0.244, v: 0.488, d: 0.324

Group # 2 -- n: 0.515, e: 0.431

Group # 3 -- a: 0.473, h: 0.618, s: 0.557

Group # 4 -- P: 0.404, C: 0.530

Group # 5 -- g: 0.515, o: 0.468, a: 0.451, s: 0.462, p: 0.622, q: 0.419

Log likelihood = -310.156

Word Order in Cicero's *Letters to Atticus*

----- Level # 4 -----

Run # 17, 38 cells:

Convergence at Iteration 6

Input 0.536

Group # 2 -- n: 0.515, e: 0.428

Group # 3 -- a: 0.471, h: 0.623, s: 0.568

Group # 4 -- P: 0.465, C: 0.511

Group # 5 -- g: 0.474, o: 0.518, a: 0.505, s: 0.478, p: 0.672, q: 0.468

Log likelihood = -316.231 Significance = 0.009

Run # 18, 37 cells:

Convergence at Iteration 8

Input 0.536

Group # 1 -- n: 0.538, p: 0.244, v: 0.490, d: 0.321

Group # 3 -- a: 0.470, h: 0.627, s: 0.563

Group # 4 -- P: 0.397, C: 0.533

Group # 5 -- g: 0.515, o: 0.468, a: 0.446, s: 0.460, p: 0.629, q: 0.425

Log likelihood = -311.013 Significance = 0.193

Run # 19, 28 cells:

Convergence at Iteration 8

Input 0.535

Group # 1 -- n: 0.539, p: 0.236, v: 0.463, d: 0.326

Group # 2 -- n: 0.518, e: 0.416

Group # 4 -- P: 0.402, C: 0.531

Group # 5 -- g: 0.515, o: 0.466, a: 0.452, s: 0.456, p: 0.633, q: 0.479

Log likelihood = -312.454 Significance = 0.101

Run # 20, 43 cells:

Convergence at Iteration 5

Input 0.536

Group # 1 -- n: 0.530, p: 0.283, v: 0.500, d: 0.367

Group # 2 -- n: 0.517, e: 0.418

Group # 3 -- a: 0.472, h: 0.622, s: 0.554

Group # 5 -- g: 0.493, o: 0.501, a: 0.490, s: 0.459, p: 0.655, q: 0.454

Log likelihood = -312.170 Significance = 0.046

Run # 21, 21 cells:

Convergence at Iteration 6

Input 0.535

Group # 1 -- n: 0.537, p: 0.256, v: 0.463, d: 0.330

Group # 2 -- n: 0.516, e: 0.425

Group # 3 -- a: 0.472, h: 0.625, s: 0.550

Group # 4 -- P: 0.408, C: 0.529

Log likelihood = -311.647 Significance = 0.703

Cut Group # 5 with factors goaspg

----- Level # 3 -----

Run # 22, 11 cells:

Convergence at Iteration 5

Input 0.535

Group # 2 -- n: 0.517, e: 0.421

Group # 3 -- a: 0.470, h: 0.630, s: 0.567

Group # 4 -- P: 0.445, C: 0.517

Log likelihood = -318.378 Significance = 0.006

A Multivariate Approach

Run # 23, 12 cells:
Convergence at Iteration 6
Input 0.535
Group # 1 -- n: 0.537, p: 0.256, v: 0.464, d: 0.327
Group # 3 -- a: 0.470, h: 0.635, s: 0.557
Group # 4 -- P: 0.400, C: 0.532
Log likelihood = -312.683 Significance = 0.160

Run # 24, 10 cells:
Convergence at Iteration 6
Input 0.534
Group # 1 -- n: 0.538, p: 0.248, v: 0.435, d: 0.331
Group # 2 -- n: 0.520, e: 0.408
Group # 4 -- P: 0.405, C: 0.530
Log likelihood = -314.202 Significance = 0.082

Run # 25, 16 cells:
Convergence at Iteration 5
Input 0.535
Group # 1 -- n: 0.531, p: 0.278, v: 0.493, d: 0.356
Group # 2 -- n: 0.519, e: 0.409
Group # 3 -- a: 0.471, h: 0.630, s: 0.548
Log likelihood = -313.918 Significance = 0.037

Cut Group # 2 with factors ne

----- Level # 2 -----

Run # 26, 6 cells:
Convergence at Iteration 4
Input 0.535
Group # 3 -- a: 0.467, h: 0.640, s: 0.573
Group # 4 -- P: 0.436, C: 0.520
Log likelihood = -319.557 Significance = 0.006

Run # 27, 5 cells:
Convergence at Iteration 6
Input 0.534
Group # 1 -- n: 0.538, p: 0.247, v: 0.433, d: 0.327
Group # 4 -- P: 0.394, C: 0.534
Log likelihood = -315.788 Significance = 0.046

Run # 28, 9 cells:
Convergence at Iteration 5
Input 0.535
Group # 1 -- n: 0.531, p: 0.281, v: 0.497, d: 0.355
Group # 3 -- a: 0.468, h: 0.643, s: 0.556
Log likelihood = -315.491 Significance = 0.018

All remaining groups significant

Groups eliminated while stepping down: 5 2
Best stepping up run: #12
Best stepping down run: #23

3. Adverbs

```
· CELL CREATION · 08/29/07 15:41:24
.....
Name of token file: Adverbs II & III.tkn
Name of condition file: Adverbs II & III.cnd
(
; Identity recode: All groups included as is.
(1)
(2)
(3)
(4)
(5)
)
```

```
Number of cells: 43
Application value(s): 10
Total no. of factors: 13
```

Group		1	0	Total	%

1 (2)		1	0		
n	N	25	96	121	20.0
	%	20.7	79.3		
p	N	35	101	136	22.5
	%	25.7	74.3		
a	N	37	180	217	35.9
	%	17.1	82.9		
v	N	2	14	16	2.6
	%	12.5	87.5		
d	N	5	110	115	19.0
	%	4.3	95.7		
Total	N	104	501	605	
	%	17.2	82.8		

2 (3)		1	0		
h	N	18	195	213	35.2
	%	8.5	91.5		
a	N	75	260	335	55.4
	%	22.4	77.6		
s	N	11	46	57	9.4
	%	19.3	80.7		
Total	N	104	501	605	
	%	17.2	82.8		

A Multivariate Approach

3 (4)		1	0		
L	N	100	489	589	97.4
	%	17.0	83.0		
G	N	4	12	16	2.6
	%	25.0	75.0		
Total	N	104	501	605	
	%	17.2	82.8		

4 (5)		1	0		
p	N	39	106	145	24.0
	%	26.9	73.1		
m	N	38	232	270	44.6
	%	14.1	85.9		
t	N	27	163	190	31.4
	%	14.2	85.8		
Total	N	104	501	605	
	%	17.2	82.8		

TOTAL	N	104	501	605	
	%	17.2	82.8		

Name of new cell file: .cel

· BINOMIAL VARBRUL · 08/29/07 15:41:29

Name of cell file: .cel

Averaging by weighting factors.
 Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:
 Convergence at Iteration 2
 Input 0.172
 Log likelihood = -277.627

----- Level # 1 -----

Run # 2, 5 cells:
 Convergence at Iteration 5
 Input 0.153
 Group # 1 -- n: 0.591, p: 0.658, a: 0.533, v: 0.442, d: 0.202
 Log likelihood = -264.895 Significance = 0.000

Word Order in Cicero's *Letters to Atticus*

Run # 3, 3 cells:
Convergence at Iteration 5
Input 0.160
Group # 2 -- h: 0.328, a: 0.603, s: 0.557
Log likelihood = -267.798 Significance = 0.000

Run # 4, 2 cells:
Convergence at Iteration 4
Input 0.172
Group # 3 -- L: 0.497, G: 0.616
Log likelihood = -277.308 Significance = 0.441

Run # 5, 3 cells:
Convergence at Iteration 4
Input 0.166
Group # 4 -- p: 0.648, m: 0.451, t: 0.454
Log likelihood = -271.791 Significance = 0.005

Add Group # 1 with factors npavd

----- Level # 2 -----

Run # 6, 14 cells:
Convergence at Iteration 5
Input 0.143
Group # 1 -- n: 0.579, p: 0.653, a: 0.537, v: 0.461, d: 0.208
Group # 2 -- h: 0.336, a: 0.590, s: 0.598
Log likelihood = -256.633 Significance = 0.000

Run # 7, 7 cells:
Convergence at Iteration 5
Input 0.153
Group # 1 -- n: 0.584, p: 0.659, a: 0.535, v: 0.444, d: 0.203
Group # 3 -- L: 0.498, G: 0.563
Log likelihood = -264.810 Significance = 0.686

Run # 8, 15 cells:
Convergence at Iteration 10
Input 0.151
Group # 1 -- n: 0.557, p: 0.600, a: 0.587, v: 0.444, d: 0.206
Group # 4 -- p: 0.603, m: 0.432, t: 0.518
Log likelihood = -262.976 Significance = 0.155

Add Group # 2 with factors has

----- Level # 3 -----

Run # 9, 18 cells:
Convergence at Iteration 5
Input 0.143
Group # 1 -- n: 0.576, p: 0.653, a: 0.537, v: 0.461, d: 0.209
Group # 2 -- h: 0.337, a: 0.590, s: 0.598
Group # 3 -- L: 0.499, G: 0.524
Log likelihood = -256.620 Significance = 0.880

A Multivariate Approach

Run # 10, 39 cells:
Convergence at Iteration 11
Input 0.141
Group # 1 -- n: 0.548, p: 0.598, a: 0.592, v: 0.449, d: 0.206
Group # 2 -- h: 0.337, a: 0.589, s: 0.605
Group # 4 -- p: 0.587, m: 0.427, t: 0.537
Log likelihood = -254.971 Significance = 0.192

No remaining groups significant

Groups selected while stepping up: 1 2
Best stepping up run: #6

Stepping down...

----- Level # 4 -----

Run # 11, 43 cells:
Convergence at Iteration 12
Input 0.141
Group # 1 -- n: 0.550, p: 0.596, a: 0.592, v: 0.448, d: 0.206
Group # 2 -- h: 0.337, a: 0.589, s: 0.606
Group # 3 -- L: 0.501, G: 0.477
Group # 4 -- p: 0.589, m: 0.427, t: 0.537
Log likelihood = -254.959

----- Level # 3 -----

Run # 12, 13 cells:
Convergence at Iteration 6
Input 0.156
Group # 2 -- h: 0.337, a: 0.593, s: 0.578
Group # 3 -- L: 0.501, G: 0.460
Group # 4 -- p: 0.629, m: 0.447, t: 0.476
Log likelihood = -263.668 Significance = 0.003

Run # 13, 17 cells:
Convergence at Iteration 11
Input 0.151
Group # 1 -- n: 0.557, p: 0.600, a: 0.587, v: 0.444, d: 0.206
Group # 3 -- L: 0.500, G: 0.504
Group # 4 -- p: 0.603, m: 0.432, t: 0.518
Log likelihood = -262.976 Significance = 0.000

Run # 14, 39 cells:
Convergence at Iteration 11
Input 0.141
Group # 1 -- n: 0.548, p: 0.598, a: 0.592, v: 0.449, d: 0.206
Group # 2 -- h: 0.337, a: 0.589, s: 0.605
Group # 4 -- p: 0.587, m: 0.427, t: 0.537
Log likelihood = -254.971 Significance = 0.882

Word Order in Cicero's *Letters to Atticus*

Run # 15, 18 cells:
Convergence at Iteration 5
Input 0.143
Group # 1 -- n: 0.576, p: 0.653, a: 0.537, v: 0.461, d: 0.209
Group # 2 -- h: 0.337, a: 0.590, s: 0.598
Group # 3 -- L: 0.499, G: 0.524
Log likelihood = -256.620 Significance = 0.193

Cut Group # 3 with factors LG

----- Level # 2 -----

Run # 16, 9 cells:
Convergence at Iteration 5
Input 0.156
Group # 2 -- h: 0.338, a: 0.593, s: 0.577
Group # 4 -- p: 0.626, m: 0.448, t: 0.477
Log likelihood = -263.706 Significance = 0.003

Run # 17, 15 cells:
Convergence at Iteration 10
Input 0.151
Group # 1 -- n: 0.557, p: 0.600, a: 0.587, v: 0.444, d: 0.206
Group # 4 -- p: 0.603, m: 0.432, t: 0.518
Log likelihood = -262.976 Significance = 0.000

Run # 18, 14 cells:
Convergence at Iteration 5
Input 0.143
Group # 1 -- n: 0.579, p: 0.653, a: 0.537, v: 0.461, d: 0.208
Group # 2 -- h: 0.336, a: 0.590, s: 0.598
Log likelihood = -256.633 Significance = 0.192

Cut Group # 4 with factors pmt

----- Level # 1 -----

Run # 19, 3 cells:
Convergence at Iteration 5
Input 0.160
Group # 2 -- h: 0.328, a: 0.603, s: 0.557
Log likelihood = -267.798 Significance = 0.000

Run # 20, 5 cells:
Convergence at Iteration 5
Input 0.153
Group # 1 -- n: 0.591, p: 0.658, a: 0.533, v: 0.442, d: 0.202
Log likelihood = -264.895 Significance = 0.000

All remaining groups significant

Groups eliminated while stepping down: 3 4
Best stepping up run: #6
Best stepping down run: #18

4. Objects

· CELL CREATION · 11/23/07 18:00:38

Name of token file: OV IV_2c.tkn
 Name of condition file: OV IV_2d.cnd

```
(
(1)
(2)
(3)
(4)
(5)
(6)
(7 (T (COL 7 T))
    (P (COL 7 P))
    (L (COL 7 s))
    (L (COL 7 G))
    (F (COL 7 F))
    (I (COL 7 I))
    (p (COL 7 p))
    (E (COL 7 E)))
)
```

Number of cells: 110
 Application value(s): 01
 Total no. of factors: 21

Group		0	1	Total	%
<hr/>					
1 (2)		0	1		
l	N	380	86	466	87.8
	%	81.5	18.5		
h	N	39	26	65	12.2
	%	60.0	40.0		
Total	N	419	112	531	
	%	78.9	21.1		
<hr/>					
2 (3)		0	1		
r	N	164	84	248	46.7
	%	66.1	33.9		
t	N	255	28	283	53.3
	%	90.1	9.9		
Total	N	419	112	531	
	%	78.9	21.1		
<hr/>					

Word Order in Cicero's *Letters to Atticus*

3 (4)		0	1		
a	N	184	79	263	49.5
	%	70.0	30.0		
s	N	110	19	129	24.3
	%	85.3	14.7		
h	N	125	14	139	26.2
	%	89.9	10.1		
Total	N	419	112	531	
	%	78.9	21.1		

4 (5)		0	1		
a	N	307	85	392	73.8
	%	78.3	21.7		
d	N	100	19	119	22.4
	%	84.0	16.0		
g	N	3	4	7	1.3
	%	42.9	57.1		
b	N	9	4	13	2.4
	%	69.2	30.8		
Total	N	419	112	531	
	%	78.9	21.1		

5 (6)		0	1		
m	N	39	22	61	11.5
	%	63.9	36.1		
l	N	217	56	273	51.4
	%	79.5	20.5		
h	N	163	34	197	37.1
	%	82.7	17.3		
Total	N	419	112	531	
	%	78.9	21.1		

6 (7)		0	1		
T	N	204	54	258	48.6
	%	79.1	20.9		
P	N	78	14	92	17.3
	%	84.8	15.2		
L	N	8	2	10	1.9
	%	80.0	20.0		
F	N	74	16	90	16.9
	%	82.2	17.8		
I	N	10	6	16	3.0
	%	62.5	37.5		

A Multivariate Approach

p	N	26	14	40	7.5
	%	65.0	35.0		
E	N	19	6	25	4.7
	%	76.0	24.0		
Total	N	419	112	531	
	%	78.9	21.1		

TOTAL	N	419	112	531	
	%	78.9	21.1		

Name of new cell file: OV IV_2d.cel

· BINOMIAL VARBRUL · 11/23/07 18:00:49

Name of cell file: OV IV_2d.cel

Averaging by weighting factors.
 Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:
 Convergence at Iteration 2
 Input 0.789
 Log likelihood = -273.559

----- Level # 1 -----

Run # 2, 2 cells:
 Convergence at Iteration 5
 Input 0.795
 Group # 1 -- l: 0.533, h: 0.280
 Log likelihood = -266.597 Significance = 0.000

Run # 3, 2 cells:
 Convergence at Iteration 5
 Input 0.816
 Group # 2 -- r: 0.306, t: 0.672
 Log likelihood = -250.101 Significance = 0.000

Run # 4, 3 cells:
 Convergence at Iteration 5
 Input 0.805
 Group # 3 -- a: 0.361, s: 0.584, h: 0.683
 Log likelihood = -260.066 Significance = 0.000

Run # 5, 4 cells:
 Convergence at Iteration 5
 Input 0.792
 Group # 4 -- a: 0.487, d: 0.580, g: 0.165, b: 0.372
 Log likelihood = -270.026 Significance = 0.074

Word Order in Cicero's *Letters to Atticus*

Run # 6, 3 cells:
Convergence at Iteration 4
Input 0.793
Group # 5 -- m: 0.317, l: 0.503, h: 0.556
Log likelihood = -269.023 Significance = 0.011

Run # 7, 7 cells:
Convergence at Iteration 5
Input 0.793
Group # 6 -- T: 0.496, P: 0.592, L: 0.510, F: 0.546, I: 0.303, p:
0.326, E: 0.452
Log likelihood = -268.981 Significance = 0.171

Add Group # 2 with factors rt

----- Level # 2 -----

Run # 8, 4 cells:
Convergence at Iteration 5
Input 0.819
Group # 1 -- l: 0.526, h: 0.324
Group # 2 -- r: 0.315, t: 0.664
Log likelihood = -246.243 Significance = 0.008

Run # 9, 6 cells:
Convergence at Iteration 4
Input 0.827
Group # 2 -- r: 0.316, t: 0.663
Group # 3 -- a: 0.374, s: 0.601, h: 0.645
Log likelihood = -240.330 Significance = 0.000

Run # 10, 8 cells:
Convergence at Iteration 5
Input 0.817
Group # 2 -- r: 0.308, t: 0.671
Group # 4 -- a: 0.499, d: 0.532, g: 0.160, b: 0.465
Log likelihood = -247.919 Significance = 0.228

Run # 11, 6 cells:
Convergence at Iteration 5
Input 0.818
Group # 2 -- r: 0.310, t: 0.668
Group # 5 -- m: 0.347, l: 0.518, h: 0.524
Log likelihood = -247.488 Significance = 0.078

Run # 12, 13 cells:
Convergence at Iteration 6
Input 0.820
Group # 2 -- r: 0.301, t: 0.677
Group # 6 -- T: 0.498, P: 0.599, L: 0.671, F: 0.497, I: 0.403, p:
0.402, E: 0.316
Log likelihood = -246.702 Significance = 0.346

Add Group # 3 with factors ash

A Multivariate Approach

----- Level # 3 -----

Run # 13, 12 cells:
Convergence at Iteration 6
Input 0.833
Group # 1 -- l: 0.536, h: 0.260
Group # 2 -- r: 0.330, t: 0.650
Group # 3 -- a: 0.351, s: 0.616, h: 0.674
Log likelihood = -233.561 Significance = 0.000

Run # 14, 20 cells:
Convergence at Iteration 5
Input 0.828
Group # 2 -- r: 0.317, t: 0.663
Group # 3 -- a: 0.375, s: 0.601, h: 0.642
Group # 4 -- a: 0.497, d: 0.531, g: 0.180, b: 0.525
Log likelihood = -238.541 Significance = 0.313

Run # 15, 18 cells:
Convergence at Iteration 4
Input 0.829
Group # 2 -- r: 0.319, t: 0.660
Group # 3 -- a: 0.376, s: 0.600, h: 0.641
Group # 5 -- m: 0.362, l: 0.522, h: 0.513
Log likelihood = -238.296 Significance = 0.139

Run # 16, 38 cells:
Convergence at Iteration 6
Input 0.831
Group # 2 -- r: 0.310, t: 0.668
Group # 3 -- a: 0.375, s: 0.595, h: 0.648
Group # 6 -- T: 0.499, P: 0.584, L: 0.703, F: 0.501, I: 0.439, p:
0.392, E: 0.333
Log likelihood = -237.349 Significance = 0.435

Add Group # 1 with factors lh

----- Level # 4 -----

Run # 17, 32 cells:
Convergence at Iteration 6
Input 0.834
Group # 1 -- l: 0.534, h: 0.273
Group # 2 -- r: 0.328, t: 0.652
Group # 3 -- a: 0.353, s: 0.615, h: 0.672
Group # 4 -- a: 0.501, d: 0.503, g: 0.291, b: 0.548
Log likelihood = -232.984 Significance = 0.764

Run # 18, 33 cells:
Convergence at Iteration 6
Input 0.835
Group # 1 -- l: 0.538, h: 0.249
Group # 2 -- r: 0.333, t: 0.648
Group # 3 -- a: 0.351, s: 0.617, h: 0.673
Group # 5 -- m: 0.355, l: 0.540, h: 0.491
Log likelihood = -231.051 Significance = 0.085

Word Order in Cicero's *Letters to Atticus*

Run # 19, 58 cells:
Convergence at Iteration 6
Input 0.836
Group # 1 -- l: 0.536, h: 0.262
Group # 2 -- r: 0.325, t: 0.655
Group # 3 -- a: 0.352, s: 0.610, h: 0.676
Group # 6 -- T: 0.513, P: 0.575, L: 0.662, F: 0.474, I: 0.440, p:
0.384, E: 0.345
Log likelihood = -230.983 Significance = 0.525

No remaining groups significant

Groups selected while stepping up: 2 3 1
Best stepping up run: #13

Stepping down...

----- Level # 6 -----

Run # 20, 110 cells:
Convergence at Iteration 18
Input 0.840
Group # 1 -- l: 0.536, h: 0.264
Group # 2 -- r: 0.325, t: 0.655
Group # 3 -- a: 0.353, s: 0.618, h: 0.669
Group # 4 -- a: 0.464, d: 0.626, g: 0.267, b: 0.561
Group # 5 -- m: 0.333, l: 0.599, h: 0.416
Group # 6 -- T: 0.480, P: 0.683, L: 0.771, F: 0.443, I: 0.366, p:
0.383, E: 0.365
Log likelihood = -225.596

----- Level # 5 -----

Run # 21, 89 cells:
Convergence at Iteration 19
Input 0.835
Group # 2 -- r: 0.312, t: 0.667
Group # 3 -- a: 0.377, s: 0.602, h: 0.638
Group # 4 -- a: 0.459, d: 0.649, g: 0.167, b: 0.547
Group # 5 -- m: 0.331, l: 0.575, h: 0.449
Group # 6 -- T: 0.494, P: 0.664, L: 0.803, F: 0.423, I: 0.364, p:
0.405, E: 0.325
Log likelihood = -231.187 Significance = 0.001

Run # 22, 83 cells:
Convergence at Iteration 20
Input 0.823
Group # 1 -- l: 0.544, h: 0.221
Group # 3 -- a: 0.339, s: 0.604, h: 0.704
Group # 4 -- a: 0.451, d: 0.663, g: 0.282, b: 0.554
Group # 5 -- m: 0.335, l: 0.591, h: 0.427
Group # 6 -- T: 0.500, P: 0.667, L: 0.659, F: 0.441, I: 0.262, p:
0.328, E: 0.463
Log likelihood = -239.940 Significance = 0.000

A Multivariate Approach

Run # 23, 65 cells:

Convergence at Iteration 19

Input 0.827

Group # 1 -- l: 0.523, h: 0.339

Group # 2 -- r: 0.311, t: 0.667

Group # 4 -- a: 0.458, d: 0.652, g: 0.206, b: 0.521

Group # 5 -- m: 0.324, l: 0.574, h: 0.453

Group # 6 -- T: 0.506, P: 0.675, L: 0.744, F: 0.396, I: 0.336, p:
0.412, E: 0.303

Log likelihood = -236.948 Significance = 0.000

Run # 24, 85 cells:

No Convergence at Iteration 20

Input 0.840

Group # 1 -- l: 0.539, h: 0.247

Group # 2 -- r: 0.323, t: 0.656

Group # 3 -- a: 0.348, s: 0.619, h: 0.675

Group # 5 -- m: 0.333, l: 0.593, h: 0.424

Group # 6 -- T: 0.452, P: 0.655, L: 0.764, F: 0.538, I: 0.469, p:
0.354, E: 0.416

Log likelihood = -227.041 Significance = 0.419

Run # 25, 80 cells:

Convergence at Iteration 20

Input 0.837

Group # 1 -- l: 0.534, h: 0.276

Group # 2 -- r: 0.326, t: 0.654

Group # 3 -- a: 0.356, s: 0.609, h: 0.671

Group # 4 -- a: 0.459, d: 0.631, g: 0.294, b: 0.618

Group # 6 -- T: 0.550, P: 0.604, L: 0.653, F: 0.368, I: 0.316, p:
0.420, E: 0.288

Log likelihood = -229.579 Significance = 0.019

Run # 26, 63 cells:

Convergence at Iteration 6

Input 0.836

Group # 1 -- l: 0.536, h: 0.264

Group # 2 -- r: 0.331, t: 0.649

Group # 3 -- a: 0.353, s: 0.616, h: 0.669

Group # 4 -- a: 0.496, d: 0.522, g: 0.265, b: 0.555

Group # 5 -- m: 0.350, l: 0.545, h: 0.485

Log likelihood = -230.237 Significance = 0.165

Cut Group # 4 with factors adgb

----- Level # 4 -----

Run # 27, 63 cells:

Convergence at Iteration 18

Input 0.833

Group # 2 -- r: 0.308, t: 0.670

Group # 3 -- a: 0.374, s: 0.601, h: 0.644

Group # 5 -- m: 0.333, l: 0.560, h: 0.470

Group # 6 -- T: 0.466, P: 0.623, L: 0.799, F: 0.529, I: 0.480, p:
0.378, E: 0.362

Log likelihood = -234.101 Significance = 0.000

Word Order in Cicero's *Letters to Atticus*

Run # 28, 60 cells:
No Convergence at Iteration 20
Input 0.822
Group # 1 -- l: 0.547, h: 0.205
Group # 3 -- a: 0.334, s: 0.606, h: 0.712
Group # 5 -- m: 0.333, l: 0.584, h: 0.437
Group # 6 -- T: 0.462, P: 0.625, L: 0.642, F: 0.573, I: 0.376, p:
0.291, E: 0.535
Log likelihood = -241.820 Significance = 0.000

Run # 29, 45 cells:
Convergence at Iteration 17
Input 0.825
Group # 1 -- l: 0.527, h: 0.317
Group # 2 -- r: 0.309, t: 0.669
Group # 5 -- m: 0.322, l: 0.562, h: 0.471
Group # 6 -- T: 0.477, P: 0.635, L: 0.739, F: 0.504, I: 0.443, p:
0.381, E: 0.350
Log likelihood = -239.383 Significance = 0.000

Run # 30, 58 cells:
Convergence at Iteration 6
Input 0.836
Group # 1 -- l: 0.536, h: 0.262
Group # 2 -- r: 0.325, t: 0.655
Group # 3 -- a: 0.352, s: 0.610, h: 0.676
Group # 6 -- T: 0.513, P: 0.575, L: 0.662, F: 0.474, I: 0.440, p:
0.384, E: 0.345
Log likelihood = -230.983 Significance = 0.020

Run # 31, 33 cells:
Convergence at Iteration 6
Input 0.835
Group # 1 -- l: 0.538, h: 0.249
Group # 2 -- r: 0.333, t: 0.648
Group # 3 -- a: 0.351, s: 0.617, h: 0.673
Group # 5 -- m: 0.355, l: 0.540, h: 0.491
Log likelihood = -231.051 Significance = 0.241

Cut Group # 6 with factors TPLFIpE

----- Level # 3 -----

Run # 32, 18 cells:
Convergence at Iteration 4
Input 0.829
Group # 2 -- r: 0.319, t: 0.660
Group # 3 -- a: 0.376, s: 0.600, h: 0.641
Group # 5 -- m: 0.362, l: 0.522, h: 0.513
Log likelihood = -238.296 Significance = 0.000

Run # 33, 18 cells:
Convergence at Iteration 7
Input 0.819
Group # 1 -- l: 0.546, h: 0.210
Group # 3 -- a: 0.336, s: 0.605, h: 0.709
Group # 5 -- m: 0.325, l: 0.528, h: 0.517
Log likelihood = -245.531 Significance = 0.000

A Multivariate Approach

Run # 34, 12 cells:
Convergence at Iteration 5
Input 0.821
Group # 1 -- l: 0.527, h: 0.315
Group # 2 -- r: 0.319, t: 0.661
Group # 5 -- m: 0.340, l: 0.532, h: 0.508
Log likelihood = -243.349 Significance = 0.000

Run # 35, 12 cells:
Convergence at Iteration 6
Input 0.833
Group # 1 -- l: 0.536, h: 0.260
Group # 2 -- r: 0.330, t: 0.650
Group # 3 -- a: 0.351, s: 0.616, h: 0.674
Log likelihood = -233.561 Significance = 0.085

Cut Group # 5 with factors mlh

----- Level # 2 -----

Run # 36, 6 cells:
Convergence at Iteration 4
Input 0.827
Group # 2 -- r: 0.316, t: 0.663
Group # 3 -- a: 0.374, s: 0.601, h: 0.645
Log likelihood = -240.330 Significance = 0.000

Run # 37, 6 cells:
Convergence at Iteration 7
Input 0.816
Group # 1 -- l: 0.545, h: 0.216
Group # 3 -- a: 0.334, s: 0.603, h: 0.714
Log likelihood = -249.031 Significance = 0.000

Run # 38, 4 cells:
Convergence at Iteration 5
Input 0.819
Group # 1 -- l: 0.526, h: 0.324
Group # 2 -- r: 0.315, t: 0.664
Log likelihood = -246.243 Significance = 0.000

All remaining groups significant

Groups eliminated while stepping down: 4 6 5
Best stepping up run: #13
Best stepping down run: #35

Texts

- Aristotle, *Ars Rhetorica*, ed. W. D. Ross, Clarendon Press, Oxford, 1959
- Caesar, G. Julius, *De Bello Gallico*, ed. H. J. Edwards, Loeb Classical Library Series, Harvard University Press, Cambridge, Mass., 1917
- Cicero, M. Tullius, *Brutus*, transl. G. L. Hendrickson, Loeb Classical Library Series, William Heinemann, London, 1939
- *De Oratore*, transl. E. W. Sutton and H. Rackham, Loeb Classical Library Series, Harvard University Press, Cambridge, Mass., 1948
- *Epistulae Ad Atticum* Vols. I-II, ed. and transl. D. R. Shackleton-Bailey, Loeb Classical Library Series, Harvard University Press, Cambridge, Mass., 1999
- *Orator*, transl. H. M. Hubbell, Loeb Classical Library Series, William Heinemann, London, 1939
- *Pro Caelio*, ed. Austin R. G., Oxford University Press, Oxford, 1933
- *Rhetorica Ad Herennium*, transl. Harry Caplan, Loeb Classical Library Series, Harvard University Press, Cambridge, Mass., 1954
- Gorgias, *Encomium of Helen*, from Η Αρχαία Σοφιστική, N. M. Skouteropoulos, ΓΝΩΣΗ, 1991
- Nepos, Cornelius, *Vitae*, ed. E.O. Winstedt, Oxford Classical Texts Series, Oxford University Press, Oxford, 1904
- Quintilianus, M. Fabius, *Institutio Oratoria* Vols I-V, ed. Russel, D. A., Loeb Classical Library Series, Harvard University Press, Cambridge, Mass., 2001
- Tacitus, *Annalium ab exessu divi Augusti libri*, ed. Henry Furneaux, Clarendon Press, Oxford, 1907
- An Intermediate Greek-English Lexicon*, ed. H. G. Liddell, Oxford University Press, Oxford, 1889
- Chambers Dictionary of Etymology*, ed. Robert K. Barnhart, Chambers Harrap Publishers, New York, 2002
- Concise Oxford Dictionary*, ed. H. W. Fowler, Oxford University Press, Oxford, 1964

References

- Arnold, Thomas, “Bradley’s Arnold” *Latin Prose Composition*, edd. Dr G. G. Bradley and J. F. Mountford, Longmans, London, 1938
- Bayley, Robert, “The Quantitative Paradigm”, in *The Handbook of Language Variation and Change*, edd. J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes, Blackwell, Oxford, 2002
- Chandler, Clive, “Caesar Stylistics: Notes”, tutorial handout, University of Cape Town, 2005
- Chomsky, Noam, *Aspects of the Theory of Syntax*, MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1965
- de Jong, Jan, “Word Order within Latin Noun Phrases”, in *Latin Linguistics and Linguistic Theory*, ed. Harm Pinkster, John Benjamins, Amsterdam, pp. 131-41, 1983
- “The Position of the Latin Subject”, in *Subordination and Other Topics in Latin*, ed. Gualtiero Calboli, John Benjamins, Amsterdam, pp. 521-40, 1989
- Del Vecchio, Tommaso, “Observations sur l’ordre des mots dans le Latin archaïque”, in *Subordination and Other Topics in Latin*, ed. Gualtiero Calboli, John Benjamins, Amsterdam, pp. 541-58, 1989
- Devine, A. M. and Laurence D. Stephens, *Latin Word Order: Structured Meaning and Variation*, Oxford University Press, Oxford, 2006
- Elerick, Charles, “Gapping, Preemptive Markedness, and Word Order in Latin”, in *Subordination and Other Topics in Latin*, ed. Gualtiero Calboli, John Benjamins, Amsterdam, pp. 559-71, 1989
- “Phenotypic linearization in Latin, word order universals, and language change”, in *Linguistic Studies on Latin*, ed. József Herman, John Benjamins, Amsterdam, pp. 67-74, 1991
- Everitt, Anthony, *Cicero*, Random House, New York, 2003
- Gries, Konrad, “Latin word order”, *The Classical Journal* 47, 1951, pp. 83-87
- Guthrie, W. K. C., *The Sophists*, Cambridge University Press, Cambridge, 1971
- Kurzová, Helena, *From Indo-European to Latin: The Evolution of a Morphosyntactic Type*, John Benjamins, Amsterdam, 1993

- Labov, William, *Sociolinguistic Patterns*, University of Pennsylvania Press, Philadelphia, 1972
- Matthews, P. H., *Oxford Concise Dictionary of Linguistics*, Oxford University Press, Oxford, 1997
- McKirahan, Richard D., *Philosophy Before Socrates*, Hackett Publishing Company, Indianapolis, 1994
- McLachlan, Kathryn, *Word Order in Cicero's Letters to Atticus*, unpublished Honours minor dissertation, 2006
- Panhuis, Dirk G. J., *The Communicative Perspective in the Sentence: A Study of Latin Word Order*, John Benjamins, Amsterdam, 1982
- Pinkster, Harm, *Latin Syntax and Semantics*, Routledge, London, 1990
- Posner, Rebecca, *The Romance Languages*, Anchor Books, New York, 1966
- Radford, Andrew, *Syntax: A minimalist introduction*, Cambridge University Press, Cambridge, 1997
- Robbins, Charles J., "Rhetoric and Latin word order", *The Classical Journal* 47, 1951, pp. 78-83
- Tagliamonte, Sali, *Analysing Sociolinguistic Variation*, Cambridge University Press, Cambridge, 2006
- Walker, Arthur T., "Some facts of Latin word-order" in *The Classical Journal* 13, 1918, pp. 644-57

heavy and more complex objects (and possibly adverbs of place as well) triggering heavy constituent shift and moving to the right of the verb.

Degree of adjacency is also an important determining factor for almost all of the constructions investigated, due to the principle of Preemptive Markedness. The closer a head and its modifier are to each other, the easier it is to see their relationship, and so the greater the likelihood of an unusual word order. On the other hand, when separated by hyperbaton, the stronger the preference is for the neutral order. One marked construction reduces the possibility of another occurring.

So Latin word order is far from irregular, although variation between neutral and marked orders is still quite common. This is no doubt due to the rich inflectional system of Classical Latin, which encodes the grammatical relationships between constituents on the words themselves, and so frees up order for other purposes. Indeed, Roman authors and orators make full use of the flexibility of the sentence, deftly juggling words so that they may emphasise some details, whilst glossing over others, or create an impression of exactitude and clear-mindedness, or simply produce euphonious and cadenced prose. Interestingly enough, the ancient grammarian Quintilian had already observed that the best hyperbaton complies with Preemptive Markedness, as well as noting that the end of sentences are focal points and that words in Latin are arranged in those positions where they will be most effective – which confirms the view that for an understanding of the grammar of Classical Latin, one cannot do better than its native speakers. Research into Latin word order has simply borne Quintilian out. Perhaps, then, we would do well to listen to what he, and the other Roman grammarians, were saying.

APPENDIX

GoldVarb Results

1. Adjectives

```
· CELL CREATION · 10/23/07 16:51:30
.....
Name of token file: Adjectives I -CPLG.tkn
Name of condition file: Adjectives I -CPLG.cnd
(
(1)
(2)
(3)
(4)
)

Number of cells: 24
Application value(s): 10
Total no. of factors: 10
```

Group		1	0	Total	%
<hr/>					
1 (2)		1	0		
p	N	118	97	215	27.7
	%	54.9	45.1		
L	N	115	261	376	48.5
	%	30.6	69.4		
d	N	28	130	158	20.4
	%	17.7	82.3		
r	N	7	10	17	2.2
	%	41.2	58.8		
G	N	8	2	10	1.3
	%	80.0	20.0		
Total	N	276	500	776	
	%	35.6	64.4		
<hr/>					
2 (3)		1	0		
n	N	229	373	602	77.6
	%	38.0	62.0		
e	N	47	127	174	22.4
	%	27.0	73.0		
Total	N	276	500	776	
	%	35.6	64.4		
<hr/>					

A Multivariate Approach

3 (4)		1	0		
a	N	225	366	591	76.2
	%	38.1	61.9		
s	N	36	88	124	16.0
	%	29.0	71.0		
h	N	15	46	61	7.9
	%	24.6	75.4		
Total	N	276	500	776	
	%	35.6	64.4		

TOTAL	N	276	500	776	
	%	35.6	64.4		

Name of new cell file: Adjectives I -CPLG.cel

· BINOMIAL VARBRUL · 10/23/07 16:51:51

.....

Name of cell file: Adjectives I -CPLG.cel

Averaging by weighting factors.

Threshold, step-up/down: 0.050001

Stepping up...

----- Level # 0 -----

Run # 1, 1 cells:

Convergence at Iteration 2

Input 0.356

Log likelihood = -505.088

----- Level # 1 -----

Run # 2, 5 cells:

Convergence at Iteration 5

Input 0.344

Group # 1 -- p: 0.699, L: 0.457, d: 0.291, r: 0.572, G: 0.884

Log likelihood = -469.848 Significance = 0.000

Run # 3, 2 cells:

Convergence at Iteration 4

Input 0.354

Group # 2 -- n: 0.528, e: 0.403

Log likelihood = -501.391 Significance = 0.008

Run # 4, 3 cells:

Convergence at Iteration 4

Input 0.354

Group # 3 -- a: 0.529, s: 0.428, h: 0.374

Log likelihood = -501.395 Significance = 0.027

Add Group # 1 with factors pLdrG